Rashmi M G

Fine structure of Gene

1. **Introduction to gene**
2. **Number and size of gene**
3. **Portion of a gene**
4. **Gene families**
5. **Classification of gene families**
6. **Homologous genes**
7. **Pseudo genes**
8. **Acquisition of new genes**
9. **Genome**
10. **Renaturation and denaturation of DNA**
11. **Highly repetitive sequences**
12. **Moderately repetitive sequences**
13. **Transposable elements**
14. **Eukaryotic chromosome**
15. **Chromosomal organization**
16. **Histone modification**

Introduction to gene

- Word gene was first used by Wilhelm Johannsen (1909)
- Functional segment of genetic material
- Locatable region of genomic sequence
- Unit of inheritance
- Associated with regulatory regions, transcribed regions and/ or other functional sequence regions
- Genes are transcribed into mRNAs which encodes proteins

Number and size of genes

- Total gene number is known for several organisms
- The bacterium with the smallest known genome *M. genitalium* has ~470 genes
- Number of genes in a eukaryote varies from 6000-40000 genes
- The number of genes in bacterial genomes is proportional to genome size
- The number of genes in a eukaryotes does not correlate with the genome size or the complexity of organism

Rashmi M G

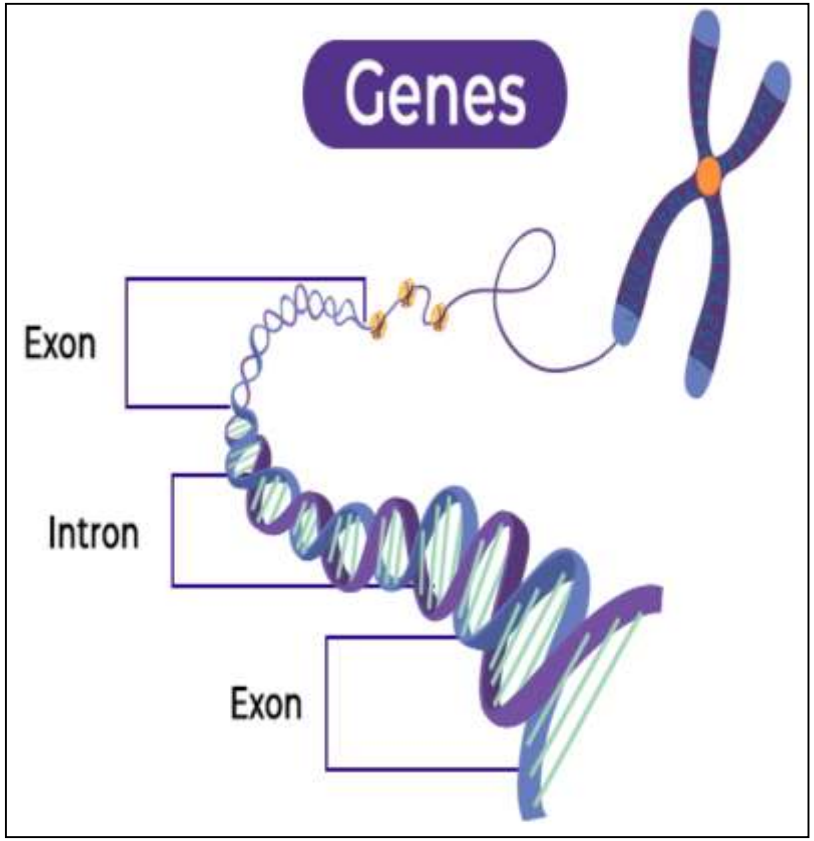
Portion of gene

Non coding DNA

- Interrupts the relatively short segments of coding DNA
- Intervening (noncoding) sequences are called **Introns**

Coding DNA

- Coding segment
- Coding Sequences are called **Exons**



Source: https://www.geeksforgeeks.org/biology/gene/

- **Introns**
- Any nucleotide sequence within a gene
- Represented in primary transcript of gene
- But not present in the final processed form
- Discovered in 1977- Philip Sharp and Richard Roberts
- Many genes like Interferons, Histone genes, Ribonuclease genes, Heat shock protein gene, G protein coupled receptors lack Introns
- Not required for gene function in eukaryotic cells
- Introns are not found in most genes of lower simple eukaryotes, found rare in genes of prokaryotes
- Some common introns: GU-AG introns, AU-AC introns, Group I introns, Group II introns, pre tRNA introns, Archaeal introns etc.

Gene families

- Set of genes with sequence homology descended from a common ancestral gene- Gene family
- Members of gene family- may be closely related/ even identical; May be clustered together/ dispersed on different chromosomes
- Gene families are highly variable in size, sequence diversity and arrangement

Classification of gene families

Based on Diversity and functions of genes within the family

Multigene families
Group of genes with related overlapping functions and arise essentially from gene duplication

Gene superfamilies
Group of related Multigene families, group of common origin with nonoverlapping functions

Simple Multigene families
Families in which all the members have identical/ nearly identical sequences
Ex. rRNA

Complex Multigene families
Families in which members are similar in sequence but different for the gene products to have distinctive properties
Ex. Mammalian globin genes

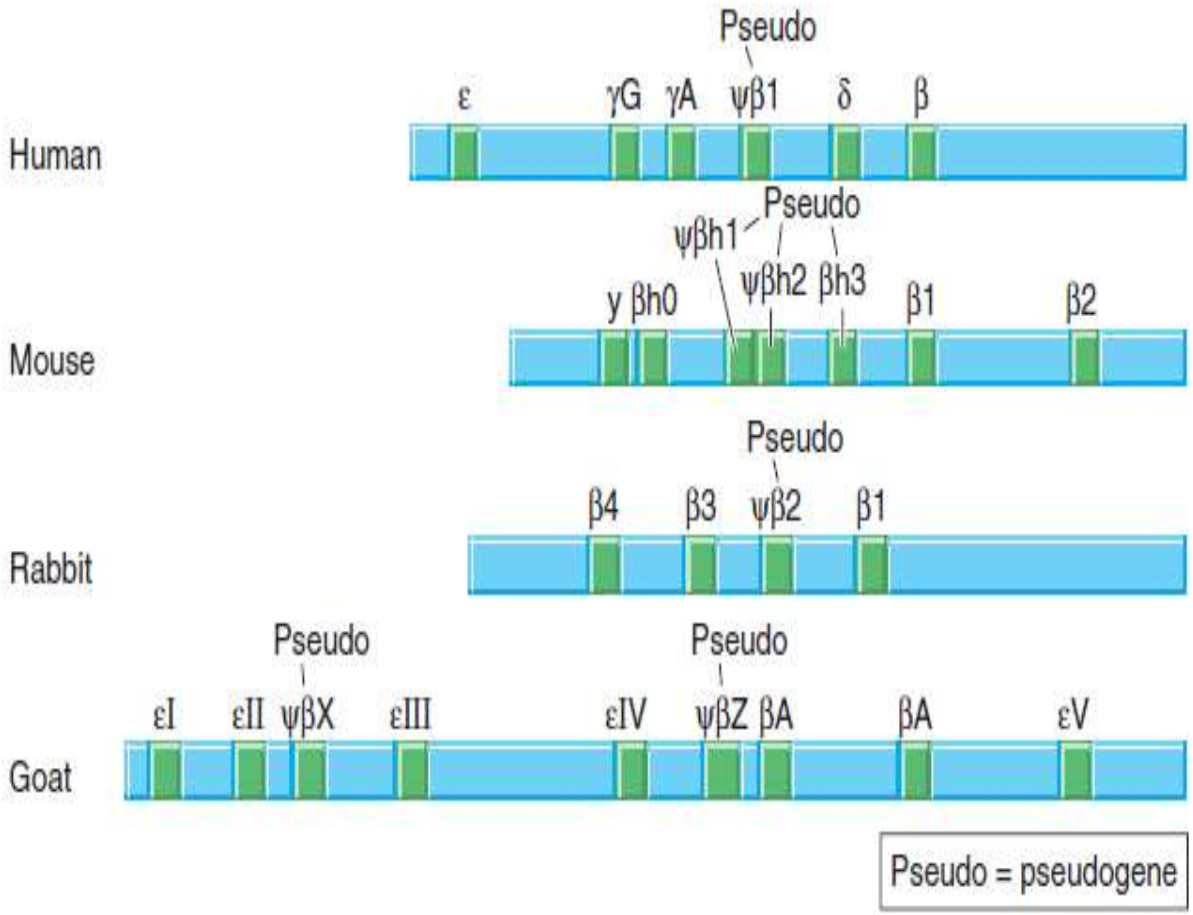
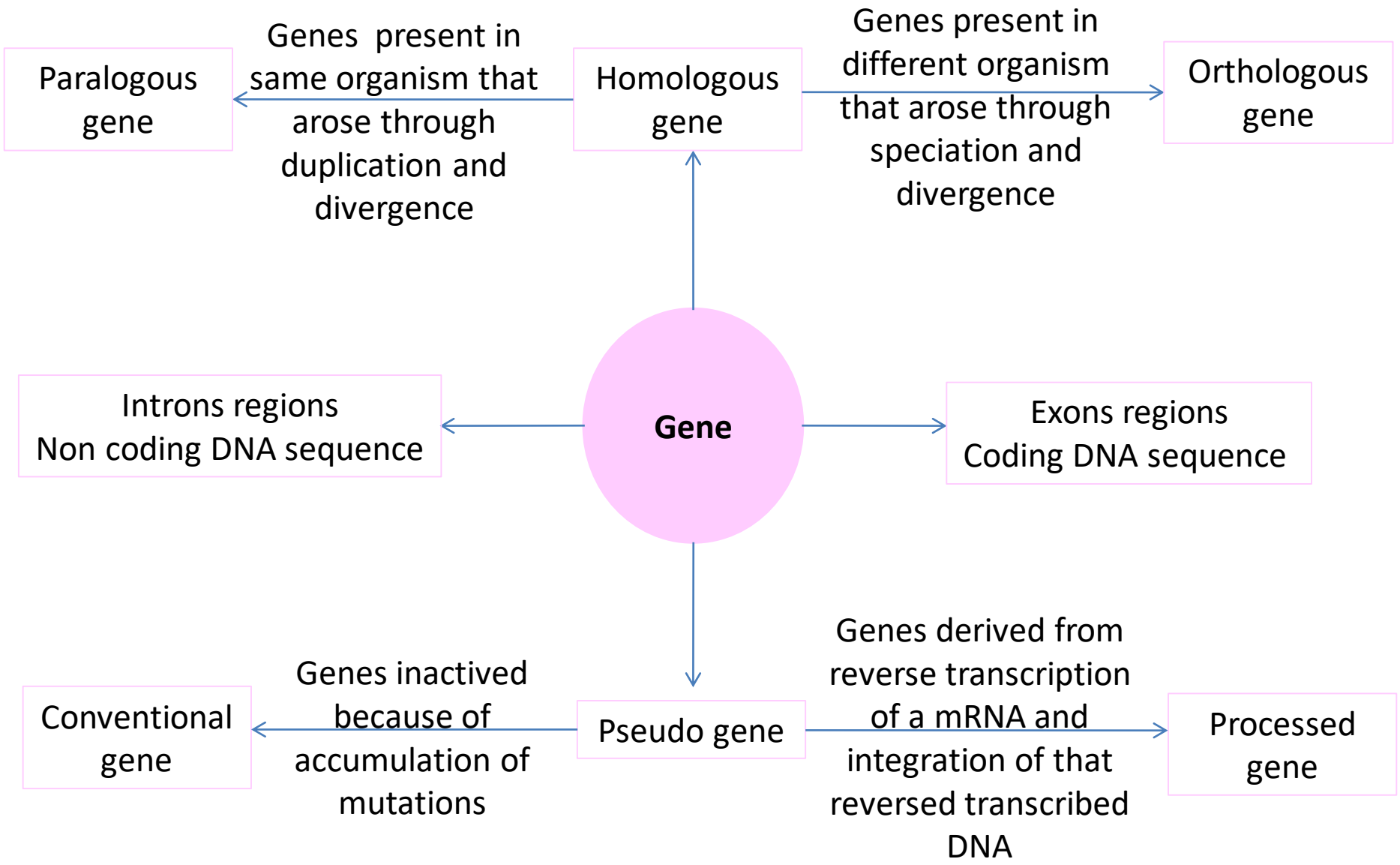


Figure 23.7

Organization of the globin gene families in several mammalian species.

Source: Peter J Russell, iGenetics A molecular approach, 3rd edition, Page no.701

Rashmi M G


Gene
Homologous gene
Genes present in different organism that arose through speciation and divergence
Orthologous gene
Paralogous gene
Genes present in same organism that arose through duplication and divergence
Introns regions
Non coding DNA sequence
Exons regions
Coding DNA sequence
Pseudo gene
Genes derived from reverse transcription of a mRNA and integration of that reversed transcribed DNA
Processed gene
Conventional gene
Genes inactived because of accumulation of mutations

Homologous genes

- Genes which share a common evolutionary ancestor
- A pair of homologous genes has sequence similarities, but do not have identical nucleotide sequences
- It is because the 2 genes undergo different random changes by mutation
- 2 Categories:
  - **Paralogous** – Homologous genes present in the same organism that arose through duplication and divergence
  - Ex. Myoglobin and –globin genes of humans
  - **Orthologous-** Homologous genes in different organism that arose through speciation and divergence
  - Ex. Myoglobin genes of humans and chimpanzees

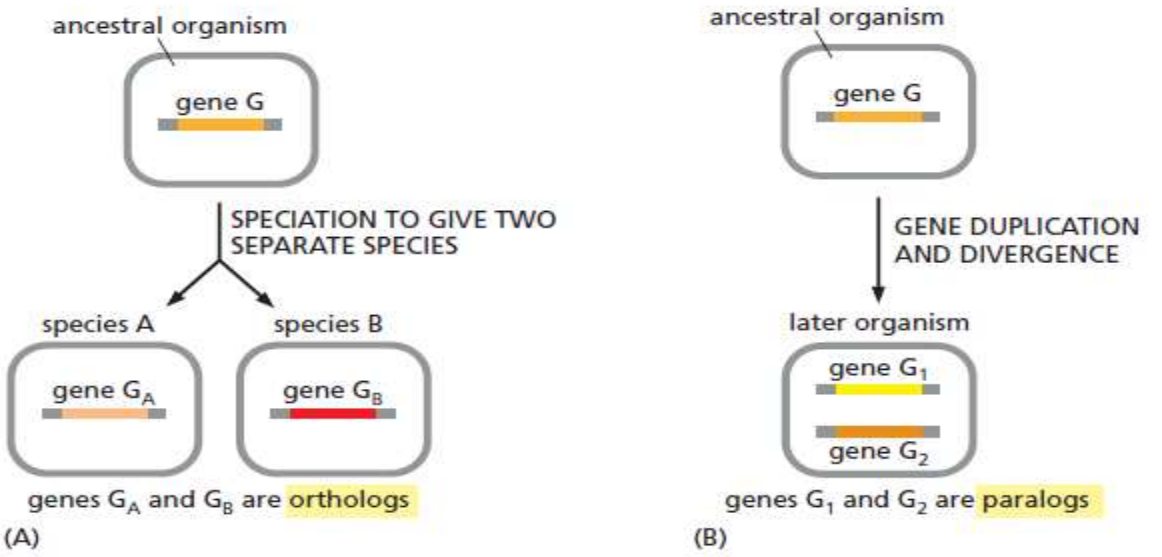


Figure 1–21 Paralogous genes and orthologous genes: two types of gene homology based on different evolutionary pathways. (A) Orthologs. (B) Paralogs.

Source: Bruce Alberts, (2015), Molecular biology of the cell, Garland Science, 6th edition, Page no. 18

Pseudo genes

- Non functional (either unexpressed/ functionless) gene copies
- Derived from functional genes
- Are the type of evolutionary relic that indicates the changing nature of the genome
- Abundance of pseudo genes- depends on rates of gene duplication and loss
- Common feature of Multigene families
- 2 Classes:
  - **Conventional (non processed) pseudo gene**- Gene that has become inactive (non functional) because of accumulation of mutations (Point, Frameshift)
  - Ex. Globin pseudo genes
  - **Processed pseudo gene**- Gene that derives from reverse transcription of a mRNA and integration of that reversed-transcribed DNA (i.e. made through mRNA retrotransposition)
  - Lack introns and promoter sequence

Acquisition of new genes

Primary driving force of evolution in all organisms

4 ways of acquisition of new genes:

1. By duplication and divergence of the existing genes in the genome
2. By acquiring genes from other species
3. Gene fusion and fission
4. De novo gene origination

1. By duplication and divergence

- May occur to intact genes or to exons
- In intact genes, Duplication- generates 2 copies of gene that are identical, but then usually the copies diverge as each accumulates different mutation
- Duplication may be result of:
- Unequal crossing over between non sister chromatids of homologous chromosomes
- Unequal sister chromatid exchange
- Replication slippage

2. By acquiring genes from other species

- Through vertical gene transfer
- Lateral gene transfer
- Gene transfer from unrelated genome (not direct inheritance from parent to offspring

3. Gene fusion and fission

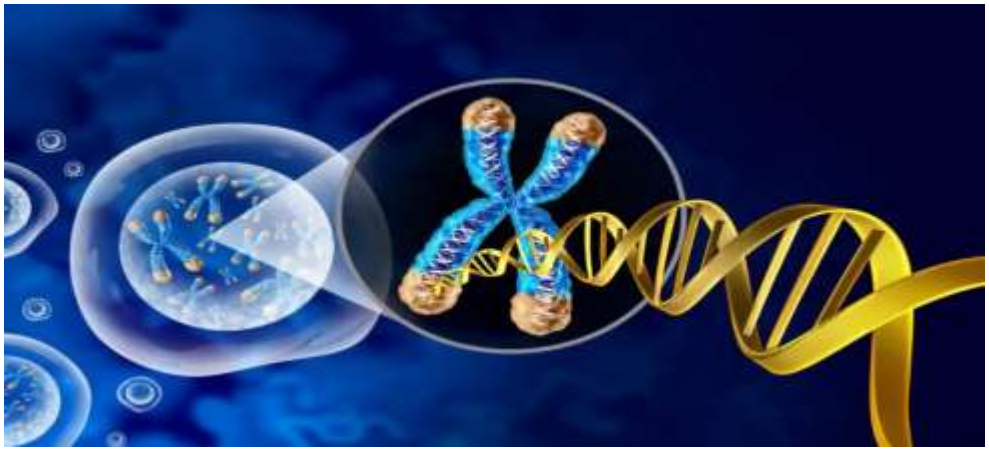
- Genes fuse (2 or more genes can become part of the same transcript) or undergo fission (single transcript can break into 2- more separate transcripts)

4. De novo gene origination

- New gene can originate de novo from noncoding regions of DNA

Genome

- **Sum total of all genetic material of an organism**
- Stores biological information
- Nature of genome may be:
  - DNA
  - RNA
- Genome:
  - Eukaryotic
  - Prokaryotic



https://www.yourgenome.org/theme/what-is-a-genome/

- Eukaryotic genome- consists of 2 distinct parts:
  1. Nuclear genome – consists of linear dsDNA (double stranded plasmid in lower eukaryotes)
  2. Organellar genome- (mitochondrial and chloroplast)
- The amount of DNA present in the genome of a species is called **C- value** which is characteristic of each species
- DNA content of the organism's genome is related to the morphological complexity of lower eukaryotes but varies extensively among the higher eukaryotes
- **Genome complexity** - the total length of different sequences of DNA
  - Can be measured through Renaturation kinetics of denatured DNA
  - **Renaturation of DNA**- complementary base-pairing again after denaturation of DNA .It depends on random collision of complementary strands

Rashmi M G

Human nuclear genome

1. Human haploid genome content is ~3.2×10^9
2. Euchromatin comprises the majority of genome ~2.9×10^9 bp
3. Human genome= 30000 genes
4. Average human gene is 27 kb long
5. Genes occupy ~25% of human genome
6. protein coding sequence ~1.5%
7. Largest gene- 2.4×10^6
8. Smallest number of exons per gene- 1
9. Largest number of exons per gene- 178
10. Largest exon size- 17106 nucleotide pairs
11. Man exon size- 145 nucleotide pairs
12. Percentage of DNA sequence in exons~1.5%
13. Percentage of DNA in high copy repetitive elements- ~50%
14. Average percentage of GC content- 41%
15. Average number of genes per chromosome- 1400

Organelle genome

Mitochondrial genome

1. Size -16.6 kbp
2. Nature of DNA molecule- circular dsDNA molecule
3. Number of DNA molecules- more than 1
4. Number of genes- 37
5. Repetitive DNA- Very little
6. % of coding DNA- 93%
7. Recombination – not evident
8. Inheritance- Maternal

Chloroplast genome

1. Nature of DNA molecule- dsDNA
2. Length- 120-190 kb
3. Multiple copies of Organellar DNA
4. In higher plants- 20-40 copies
5. Codes for all the rRNA and tRNA and
6. 50-100 proteins

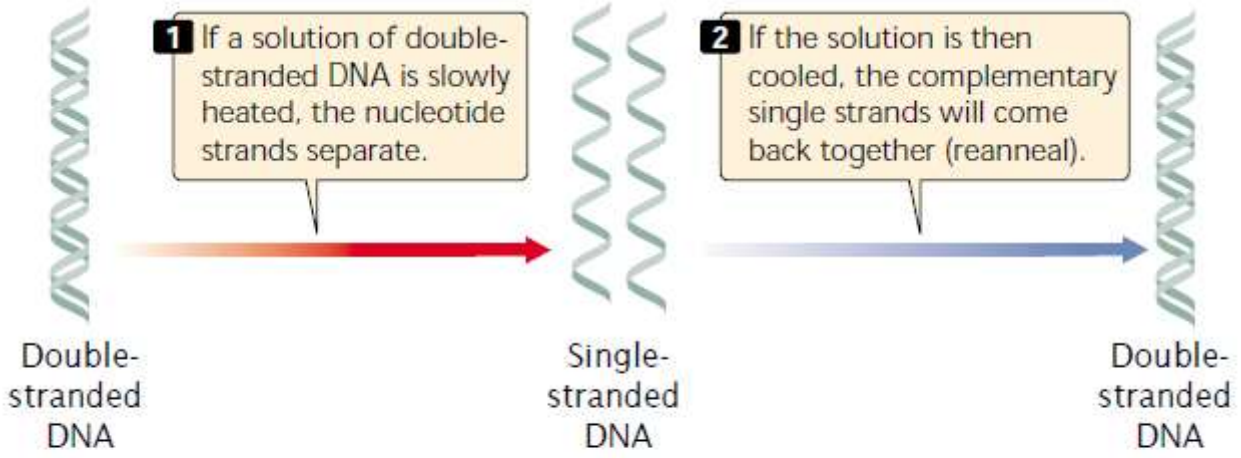
E coli genome

1. Single main chromosome and plasmids
2. Main chromosome- circular dsDNA
3. 4288 actual and proposed gene coding sequences
4. 88% of genome- encodes proteins/ RNA
5. 11% of genome – gene regulatory functions
6. <1% repetitive DNA sequences
7. Average distance between E coli- 12bp

Yeast S cerevisiae genome

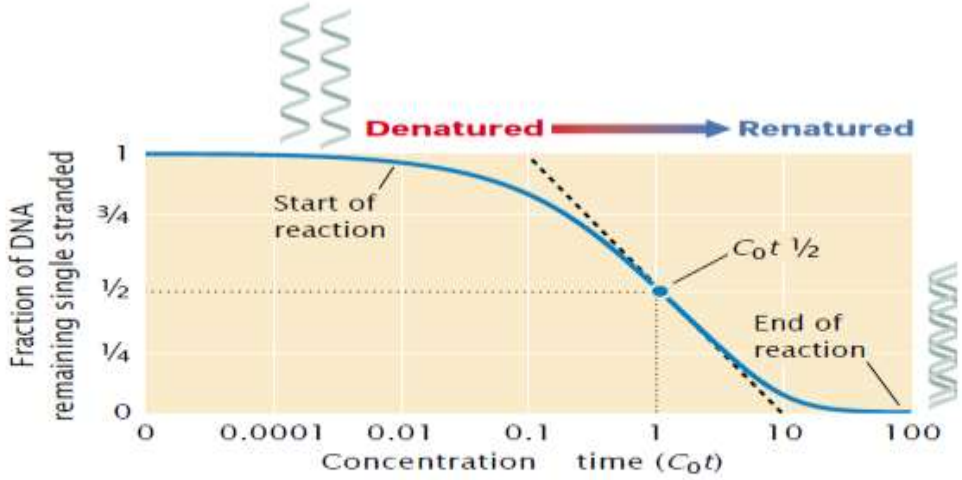
1. 16 linear chromosomes
2. 12068 kb DNA
3. 5885 potential protein coding genes, 45S RNA coding genes
4. 70% of genome- devoted to protein coding sequences
5. 4% of about 6000 yeast genes have introns

Renaturation and denaturation of DNA


1 If a solution of double-stranded DNA is slowly heated, the nucleotide strands separate.
2 If the solution is then cooled, the complementary single strands will come back together (reanneal).
Double-stranded DNA
Single-stranded DNA
Double-stranded DNA

11.13 The slow heating of DNA causes the two strands to separate (denature).

- The rate of Renaturation can be followed spectrophotometrically by UV absorbance decrease as duplex DNA is formed
- The $cot_{½}$ of a reaction described as kinetic complexity
- It indicates the total length of different sequences that are present
- This reveals sequences in DNA



11.14 A C_0t curve represents the fraction of DNA remaining single stranded in a renaturation reaction, plotted as a function of DNA concentration × time (C_0t). This graph is a typical C_0t curve for a prokaryotic organism.

Rashmi M G

Reassociation kinetics identifies 2 general types of genomic sequences:

1. **Non repetitive DNA**
2. **Repetitive DNA**

Genomic Sequences

1. Non repetitive DNA

- Sequence –unique
- Only one copy in haploid genome
- Show less rapid Renaturation rates
- Present in Prokaryotes, Lower eukaryotes

2. Repetitive DNA

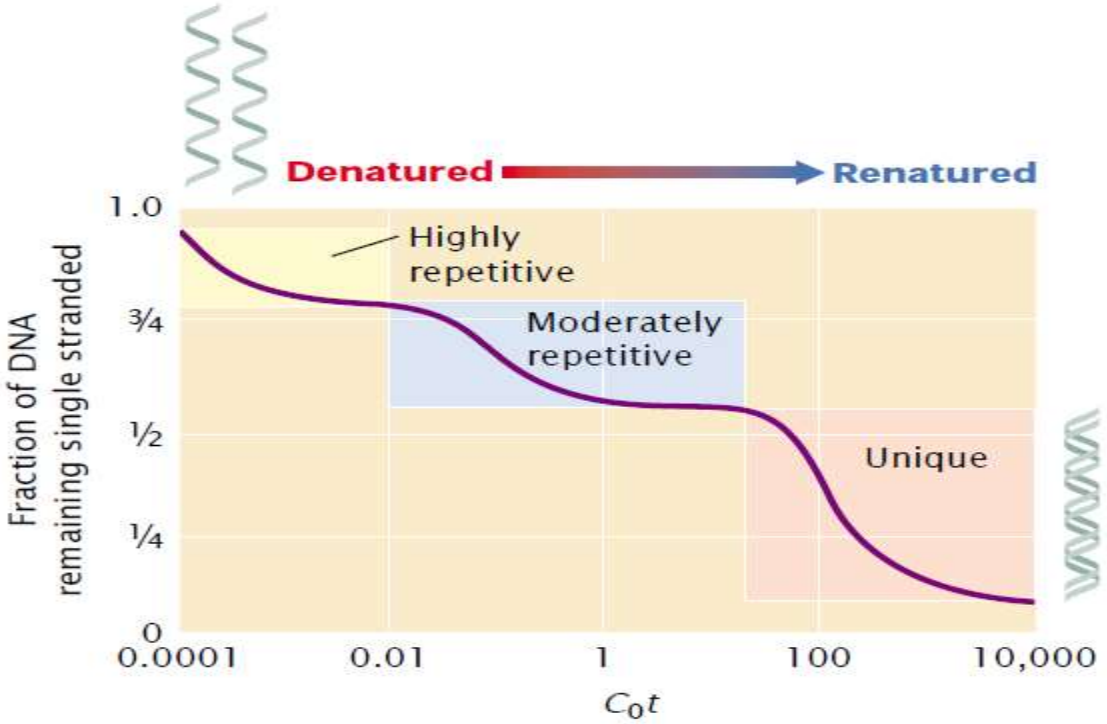
- Present more than one copy in haploid genome
- Show more rapid Renaturation rate
- Eukaryotes have half of their DNA occupied by this sequence

Moderately repetitive DNA sequences

- Relatively consisting of short sequences
- Repeated typically 10-1000x in genome
- Dispersed throughout genome
- Significant fraction of moderately repetitive DNA consist of transposable elements
- Ex. **Transposable elements**

Highly repetitive DNA sequence

- Consist of very short sequence (<1000bp)
- Repeated many times tandemly in large cluster
- Rapid rate of Renaturation \component that renatures most rapidly in eukaryotic genome
- Called simple sequence DNA
- Ex. **Satellites**



Source: Benjamin A Pierce, Genetics, A conceptual approach, Page no. 300,301,302

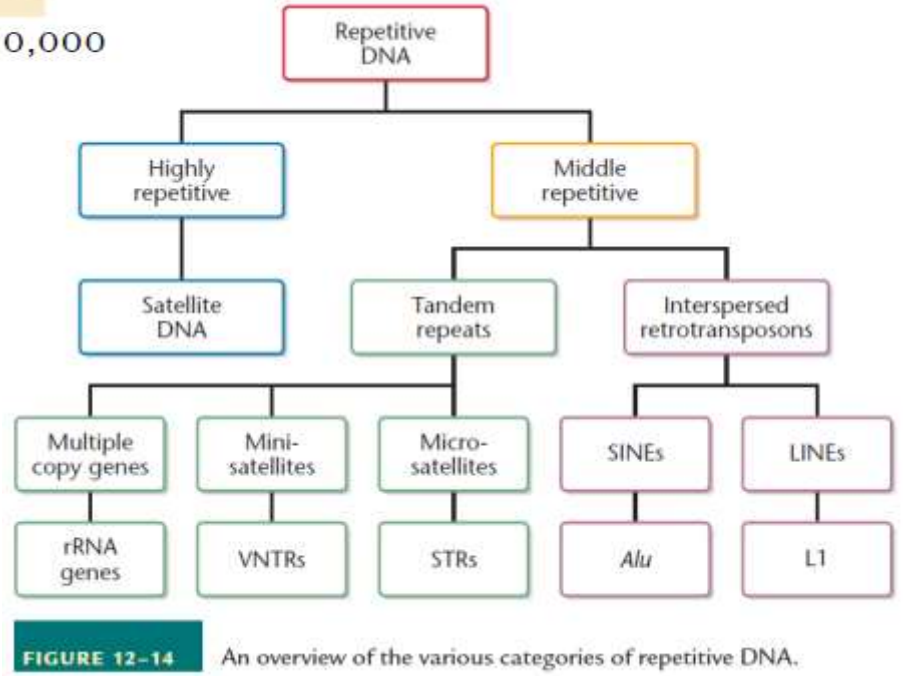


FIGURE 12-14 An overview of the various categories of repetitive DNA.

Source: Klug cummins, Spencer, Palladino (2012), Concepts of Genetics, Pearson Publication, 10th Edition, Page no.306

Highly repetitive sequences

- Satellites
- Microsatellites
- Minisatellites

Tandem repetition of short sequence → Create fraction with distinct base composition from genome average → Difference in base composition create difference in buoyant density

↓

Simple sequence DNA can often be separated from bulk of chromosomal DNA by shear degradation → CsCl density gradient centrifugation → Separates DNA on the basis of their difference in Buoyant density

↓

Formation of Satellites to main DNA band → Sequence of these DNA **Satellite DNA**

- Satellite DNA is essentially synonymous with simple sequence DNA
- Ex. When mouse DNA is centrifuged on density gradient, 2 types of bands may be distinguished- main band and a satellite
- Main band contains 92% of genome and is centered on a buoyant density of 1.701g/cm^3
- The smaller peak represents 8% of genome and has a distinct buoyant density of 1.690g/cm^3

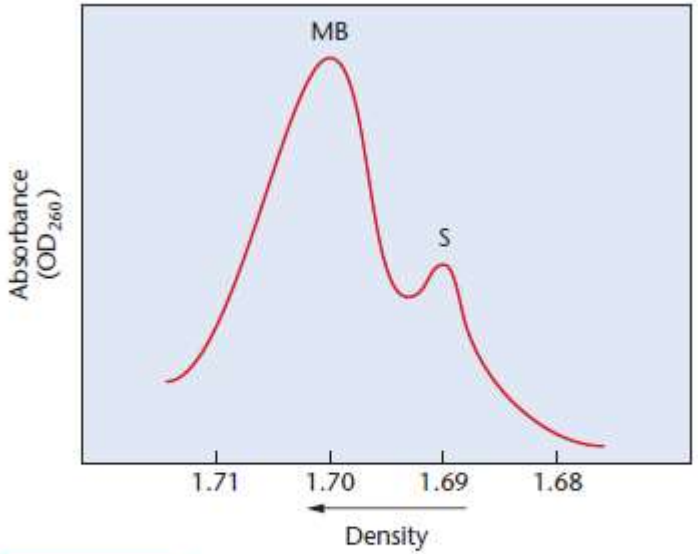
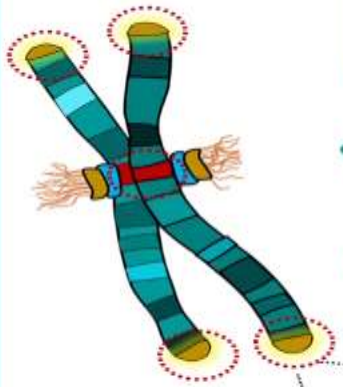


FIGURE 12–15 Separation of main-band (MB) and satellite (S) DNA from the mouse by using ultracentrifugation in a CsCl gradient.

Source: Klug cummins, Spencer, Palladino (2012), Concepts of Genetics, Pearson Publication, 10th Edition, Page no.307

Rashmi M G

Satellite DNA



Where can we find satellite dna ?

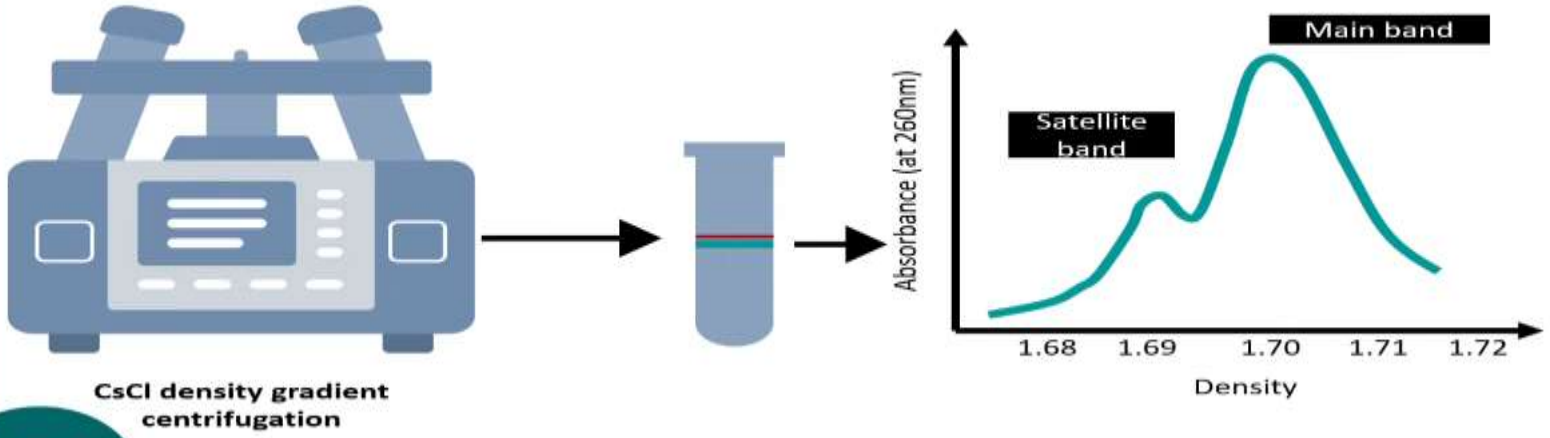
- mostly in centromeric and subtelomeric regions in the chromosomes but also at intercalary positions

What are the features of satellite dna ?

- Satellite DNA (satDNA) is the highly repetitive DNA consisting of short sequences repeated a large number of times.
- It carries a variable AT-rich repeat unit that often forms arrays up to 100 Mb.

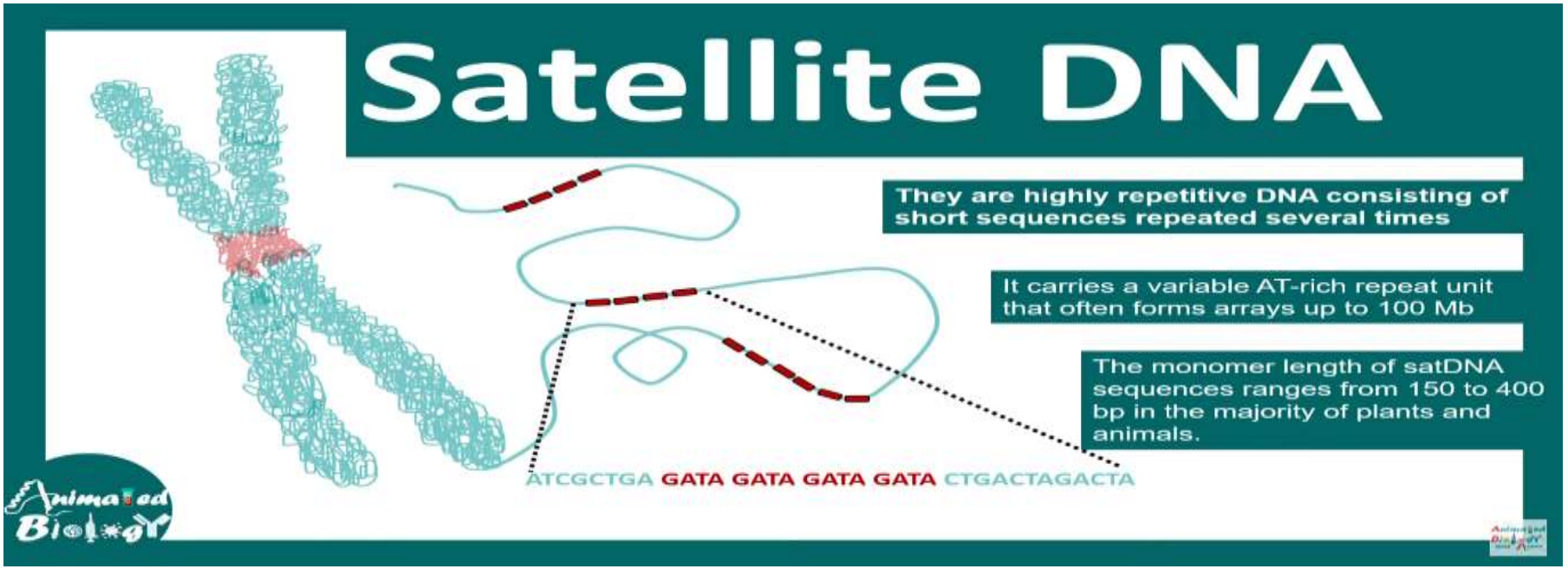
ATCGCTGA GATA GATA GATA GATA GATA GATA GATA GATA GATA GATA CTGACTAGACTA

Satellite DNA forms a different Band in CsCl density gradient centrifugation



CsCl density gradient centrifugation

Source: https://www.youtube.com/watch?v=228VEIcN56s


Satellite DNA
They are highly repetitive DNA consisting of short sequences repeated several times
It carries a variable AT-rich repeat unit that often forms arrays up to 100 Mb
The monomer length of satDNA sequences ranges from 150 to 400 bp in the majority of plants and animals.
ATCGCTGA GATA GATA GATA GATA CTGACTAGACTA

Source: https://www.youtube.com/watch?v=228VEIcN56s

Satellites

- Present in many eukaryotic genomes
- May be either heavier/ lighter than the main band
- Ex. Human satellite DNA, the alphoid DNA, repeats are found in the Centromere regions of chromosomes
- Species specific
- A single genome can contain several different types of satellite DNA, each with a different repeat unit
- These units being anything from <5 to 200bp
- Satellite DNA sequences generally are not transcribed and are located most often in the heterochromatin associated with the centromeric regions of chromosomes

Minisatellites

- Form clusters upto 20kb in length
- 20-50 repeat units
- Each containing about 10-100 base pairs
- Ex. Telomeric DNA in humans

Microsatellites

- Clusters are shorter
- <150 bp
- Repeat units- 13 bp or less
- Typical microsatellites consists of a 1-,2-,3- or 4-bp unit repeated 10 to 20 times
- These types of sequences are also called VNTRs (Variable Number Tandem Repeats) due to variation in the number of repeating units at a particular locus

Class	Size of repeat unit (bp)	Location
Satellite DNA (100kb to several Mb)	5-171	Mainly at Centromere
Alpha satellite	171	Centromeric region
Beta satellite	68	Peri-centric region of chromosome
Minisatellites DNA (upto 20 kb)	9-64	At or close to telomere
Microsatellite DNA (less than 150 bp)	2-13	Dispersed throughout all chromosome

Moderately repetitive sequences

- Transposable elements

Moderately repetitive sequences

- Consists of relatively short sequences that are repeated typically 10-1000x in the genome
- These sequences are dispersed throughout the genome
- Significant fraction of the moderately repetitive DNA consists of **transposable elements**

Transposable elements

- Short sequence of DNA
- Have the ability to move in genome
- Make duplicate copies of themselves during transposition (sequence copied and inserted into new site)
- Moderately repeated sequence interspersed at multiple sites throughout genome
- Types:
  - On the basis of mechanism of movement-Transposons and Retrotransposons
  - On the basis of self sufficiency- Autonomous and non-autonomous transposable elements
- Also called- **Selfish DNA, Jumping genes, Insertion sequences, Cassette, Mobile elements**
- Transposition- lead to mutation, major source of genome diversity and change
- Interrupts coding sequence and inactivates expression of gene
- Elements contain transcription/ translation termination signals that blocks expression of other genes downstream of insertion site
- 1st transposons- Discovered in Maize (Zea mays) by Barbara Mc Clintock in 1948

Transposable elements

Based on mechanism of movements

Transposons (Type II

transpose to new sites directly as DNA

Retrotransposons (Type I)

First transcribed into an RNA copy of the element, which then reverse-transcribed into DNA

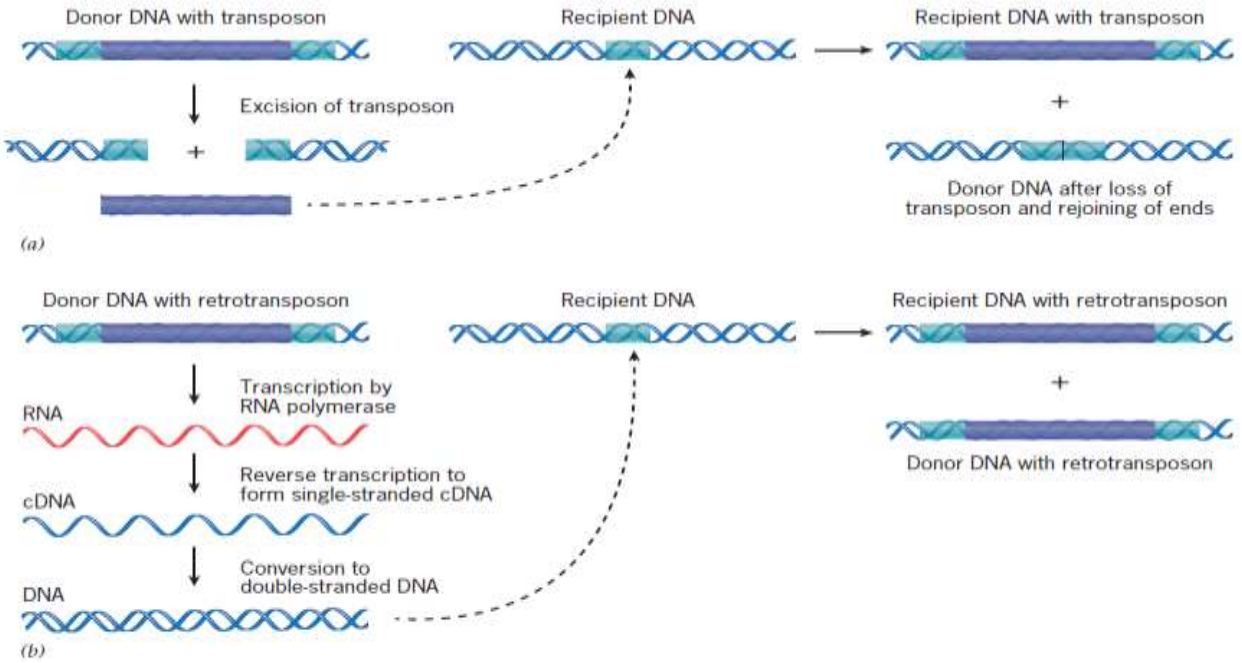


FIGURE 10.27 Schematic pathways in the movement of transposable elements. (*a*) DNA transposons move by a cut-and-paste pathway whose mechanism is depicted in Figure 10.26. Approximately 3 percent of the human genome consists of DNA transposons, none of which are capable of transposition (i.e., all are relics left in the genome as a result of ancestral activity). (*b*) Retrotransposons move by a copy-and-paste pathway. The steps involved in retrotransposition take place both in the nucleus and cytoplasm and require numerous proteins, including those of the host. More than 40 percent of the human genome consists of retrotransposons, only a few of which (e.g., 40–100) are thought to be capable of transposition. More than one mechanism of retrotransposition is known.

Source: Janet Iwasa, Wallace Marshall (2016), Karp's cell and molecular biology, 8th edition, Page no.393

Transposable Genetic Element	Structure	Genes Encoded	Transposition	Examples
Class I	Short, terminal inverted repeats; short flanking direct repeats at target site	Transposase gene (and sometimes others)	By DNA intermediate (replicative or nonreplicative)	*IS1 (E. coli)* *Tn3 (E. coli)* *Ac, Ds* (maize) *P* elements *(Drosophila)*
Class II (retrotransposon)	Long, terminal direct repeats; short flanking direct repeats at target site	Reverse transcriptase gene (and sometimes others)	By RNA intermediate	*Ty* (yeast) *copia (Drosophila)* *Alu* (human)

Source: Benjamin A Pierce, Genetics, A conceptual approach, Page no. 315

Transposable elements

Based on self sufficiency

Autonomous transposons

Has everything they need to promote their own transposition i.e. functional gene that encodes an active transposase enzyme and terminal inverted repeat

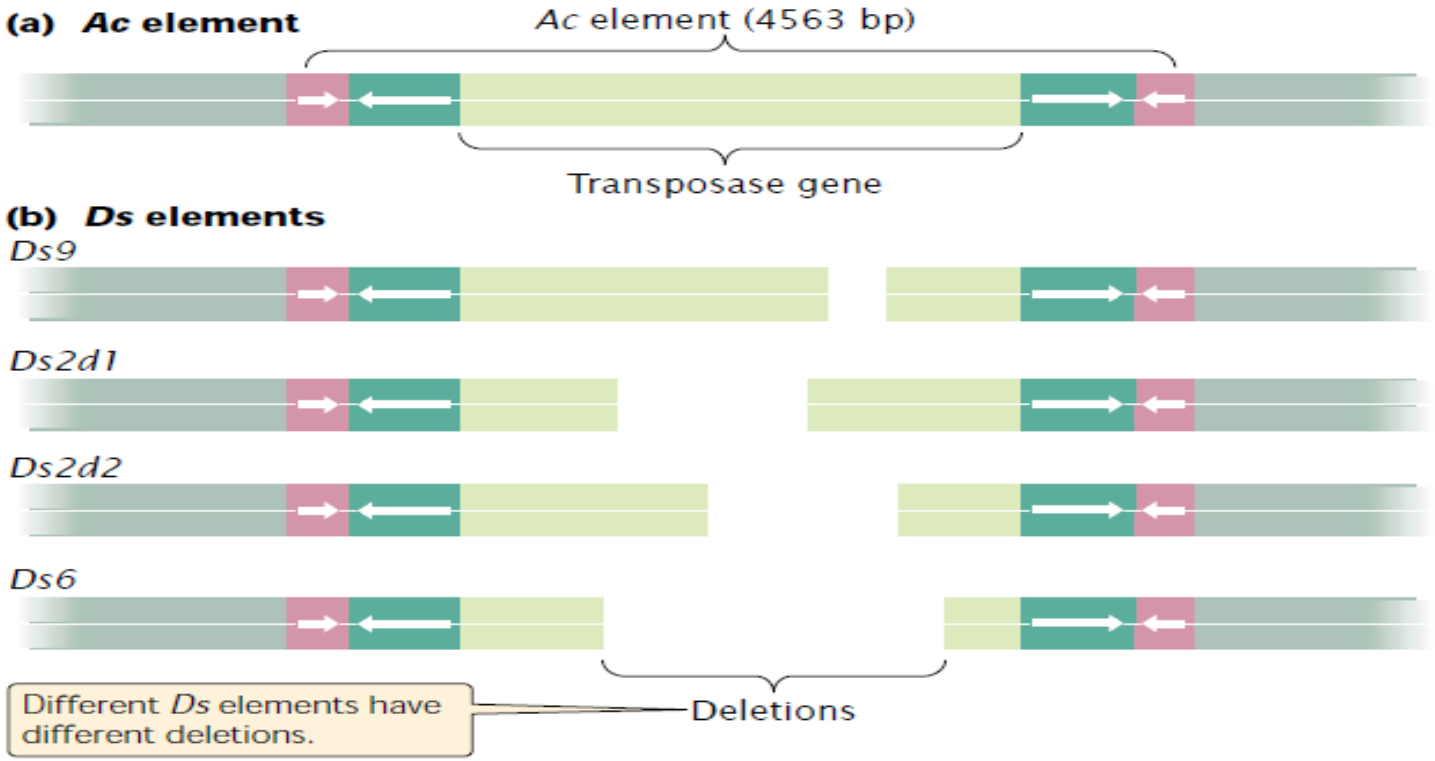
Ex. AC element

Non autonomous transposons

Carry only the terminal inverted repeats but not encode functional transposase

They require transposase encoded by autonomous transposons for transposition

Ex. DS element



Source: Benjamin A Pierce, Genetics, A conceptual approach, Page no. 311

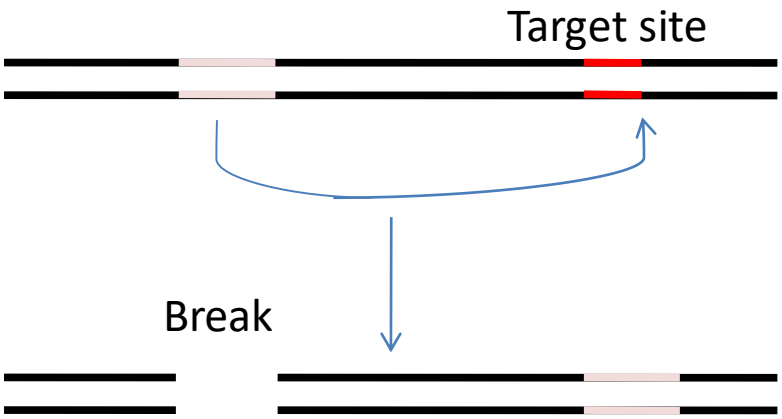
Mechanism of transposition

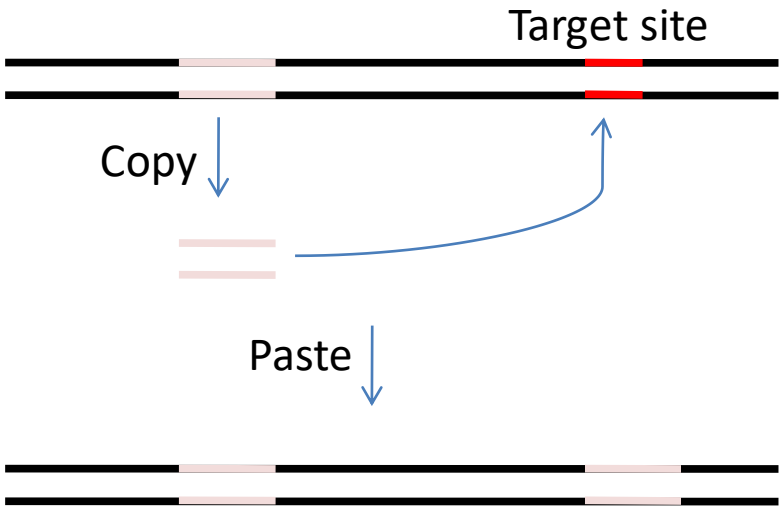
Conservative process

- Cut and paste mechanism
- Excision of sequence from its original position
- Reinsertion elsewhere
- Change in position of transposons in genome without increasing the copy number
- Requires 2 elements: Transposase and inverted repeats
- Transposons using this mechanism are known as- **Simple transposons**

Replicative process

- Copy and paste mechanism
- Creation of 2nd copy
- Original transposons remain in place
- Copy is inserted at the new position
- Transposons using this mechanism are known as- **Complex transposons**





Transposable elements

Prokaryotic transposons

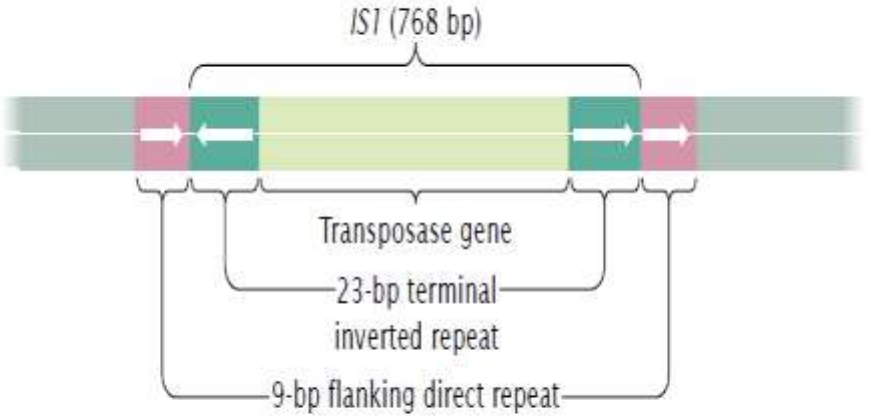
- Transposons present in prokaryotic organisms
- Ex. IS elements (simple transposons); Tn9, Tn 10 (Composite elements)

Eukaryotic transposons

- Transposons present in eukaryotic organisms
- Ex. Retrotransposons (LTR and Non LTR); P elements (Drosophila); Controlling elements (Maize)

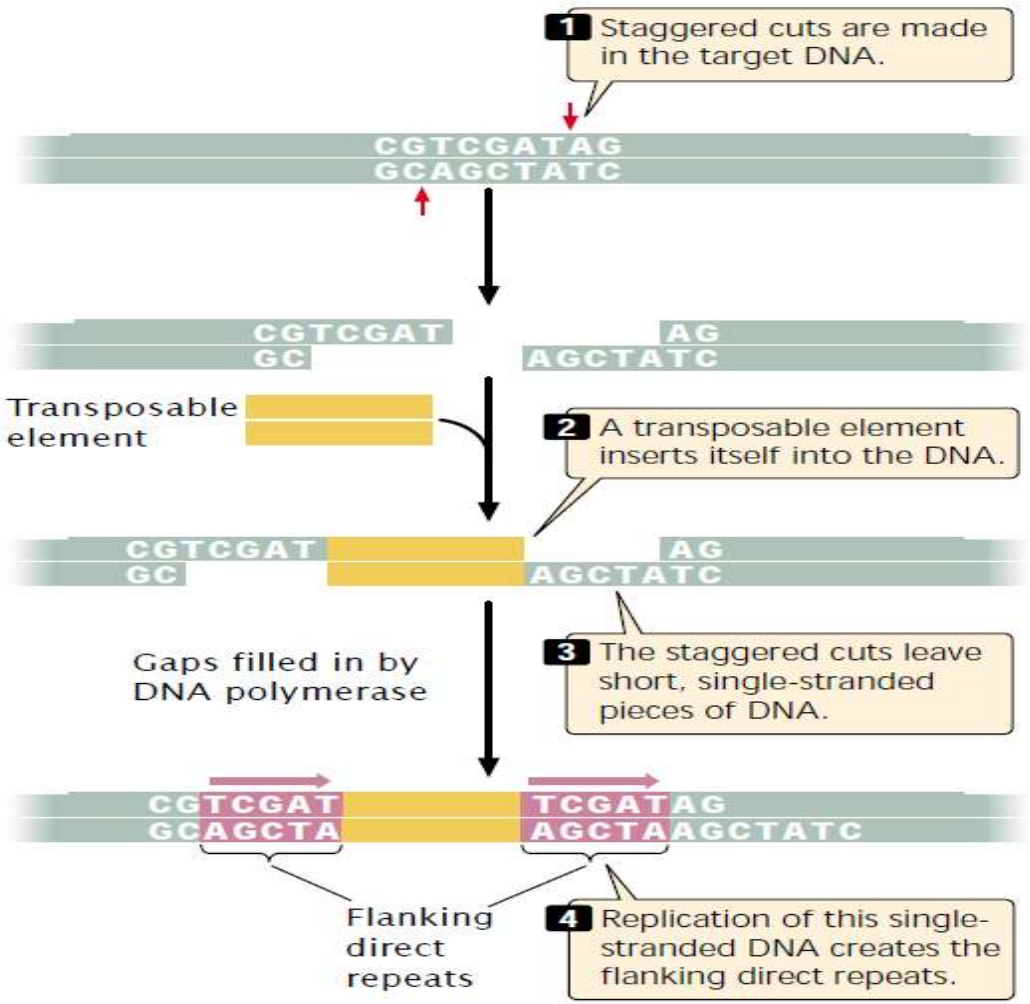
Bacterial transposons

- Transposable elements are of Type II (transpose directly to DNA)
- Produce short direct repeats at site of insertion which flank the mobile element
- Length of direct repeats depend on the type of mobile element
- IS element:
  - Insertion sequence element
  - Most common transposons present in bacterial genome
  - General structure:
    - Inverted repeat- 50 bp present at each end of an insertion sequence
    - Protein coding region- between inverted repeats, codes for one/ 2 enzymes required for transposition
    - Short direct repeats -5-10 bp adjacent to both ends of inserted elements



11.22 Insertion sequences are simple transposable elements found in bacteria.

Source: Benjamin A Pierce, Genetics, A conceptual approach, Page no. 308



11.16 Flanking direct repeats are generated when a transposable element inserts into DNA.

Source: Benjamin A Pierce, Genetics, A conceptual approach, Page no. 303

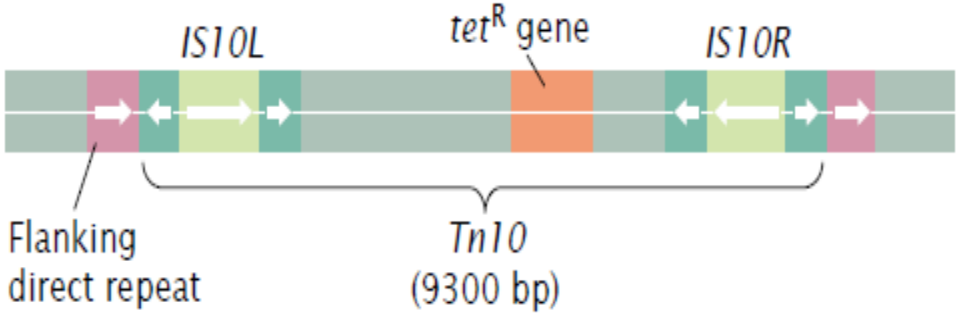
Transposition of IS element

Transposase performs 3 functions:

1. precisely excises the IS element from donor DNA
2. Makes staggered cuts in target DNA
3. Ligates the 3’ termini of IS element to 5’ ends of the cut donor DNA

Composite transposons

- In addition to IS elements, Bacteria contains composite transposons
- Larger than IS elements
- Contain 1/ more protein coding genes in addition to those required for transposition
- Consists of 2 IS elements surrounding a central block of protein coding genes
- Many of well known composite transposons carry genes for antibiotic resistance
- Ex. Tn9 (Chloramphenicol resistance); Tn10 (Tetracycline resistance)



11.23 ***Tn10*** **is a composite transposon in bacteria.**

Source: Benjamin A Pierce, Genetics, A conceptual approach, Page no. 309

Eukaryotic transposons

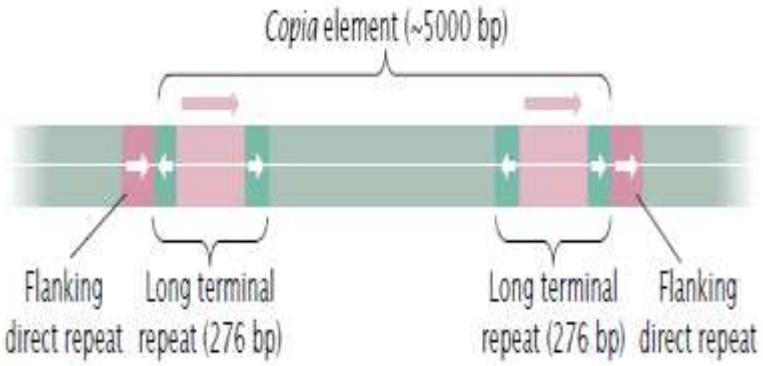
In higher eukaryotes- Retrotransposons are found

P elements:

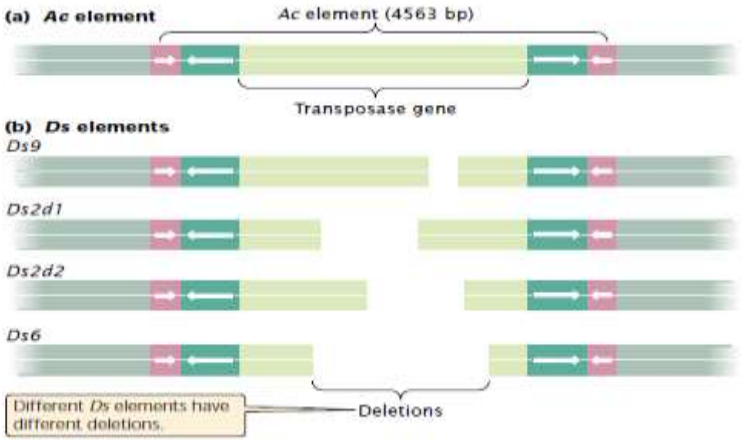
- Found in Drosophila
- 2907 bp long
- 31 bp inverted repeat at each end
- Element encodes- transposase
- Do not utilize an RNA intermediate during transposition
- Can insert at many different positions in the Drosophila chromosome
- Causative agents for hybrid dysgenesis (series of defects including mutations, chromosomal aberrations, distorted segregation at meiosis and sterility)

Controlling elements:

- Found in Maize (Discovered by Barbara Mc Clintock)
- 2 classes of controlling elements: Dissociation (Ds) and Activator (Ac)
- Ac – Autonomous transposons
- Ds- Non autonomous transposons


Copia element (~5000 bp)
Flanking direct repeat
Long terminal repeat (276 bp)
Long terminal repeat (276 bp)
Flanking direct repeat

11.30 ***Copia* is a transposable element of *Drosophila*.**



Source: Benjamin A Pierce, Genetics, A conceptual approach, Page no. 308,311

Retrotransposons

Transposable elements transpose through RNA intermediate by utilizing reverse transcriptase

Found in eukaryotic organisms

Based on presence/ absence of direct repeats at ends

2 types:

LTR Retrotransposons (viral transposons)

- Similar to retroviral DNA
- Flanked by 250-660 bp LTR
- Encode reverse transcriptase and integrase
- Move in genome by being transcribed into RNA which then undergo Reverse transcription and integration into host cell chromosome
- Found in yeast (Ty elements), Drosophila (copia elements)

Non LTR (Nonviral Retrotransposons)

- Lack LTR
- Have A/T rich stretch at one end
- In vertebrates SINES- (Short interspersed nuclear elements) and LINES (Long interspersed nuclear elements)
- Ex. L1, B1

Viral transposons

- E coli bacteriophage Mu is an unusual phage
- It is both a virus and a transposons
- It is temperate phage so it has both lytic and lysogenic cycles
- When DNA of Mu enters E coli, its host, it integrates into the host genome
- It integrates randomly into the host genome by a transpositional mechanism
- When Mu replicates, it does so by replicative transposition events in which each new copy of Mu is inserted somewhere else in the genome
- DNA of Mu is never found replicating as an independent molecule, free of the chromosome
- The segments of DNA packages into the virus particle contain whole Mu genome plus small stretches of host DNA at the ends
- Thus whole of the Mu genome is a transposons that happens to have genes that package its DNA as well as genes that regulate its transposition

Rashmi M G

Structure of Eukaryotic chromosome

Eukaryotic chromatin and chromosome

Chromatin – organized structure of DNA and protein found in nucleus of eukaryotic cell

In G˚ phase- chromatin contain single dsDNA (CONDENSED FORM)

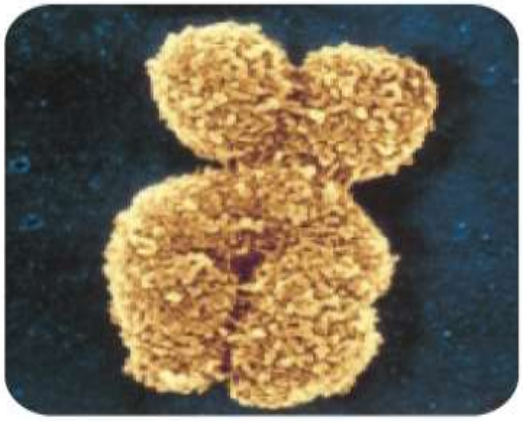
Chromatin v/s chromosome:

Chromatin – less condensed, extended DNA

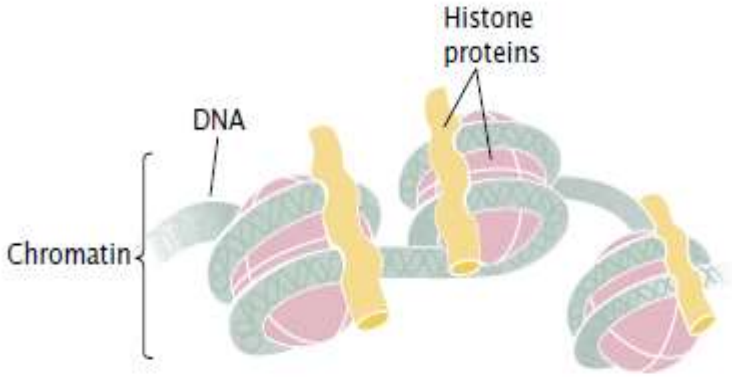
Chromosome- highly condensed, packed DNA; Chroma- color; Some- body, strongly stained by particular dyes

In interphase nuclei, Chromatin + nuclear matrix (proteinaceous structure)

DNA sequence attached to nuclear matrix at Matrix attachment region (MAR)



1.12 **Genes are carried on chromosomes.**
(Biophoto Associates/Science Source/Photo Researchers.)



2.2 **In eukaryotic cells, DNA is complexed to histone proteins to form chromatin.**

Source: Benjamin A Pierce, Genetics, A conceptual approach, Page no. 14,18

Rashmi M G

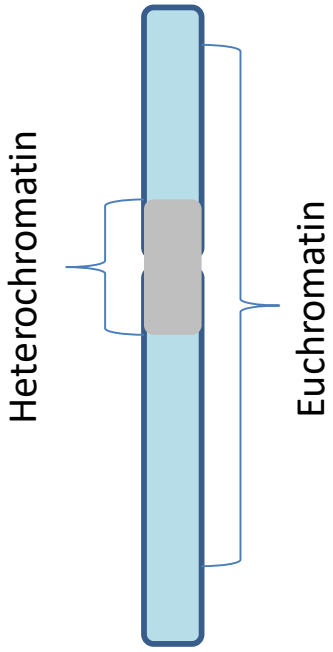
Chromatin

Euchromatin

- Light staining, less condensed portion
- Less condensed portion
- Region of transcriptionally active
- Contain most of protein coding genes
- Have an open accessible chromatin structure characterized by acetylation and methylation at H3K4, H3K9

Heterochromatin

- Dark staining highly condensed portion
- Remains condensed
- Condensed state during interphase-transcriptionally silent



Constitutive

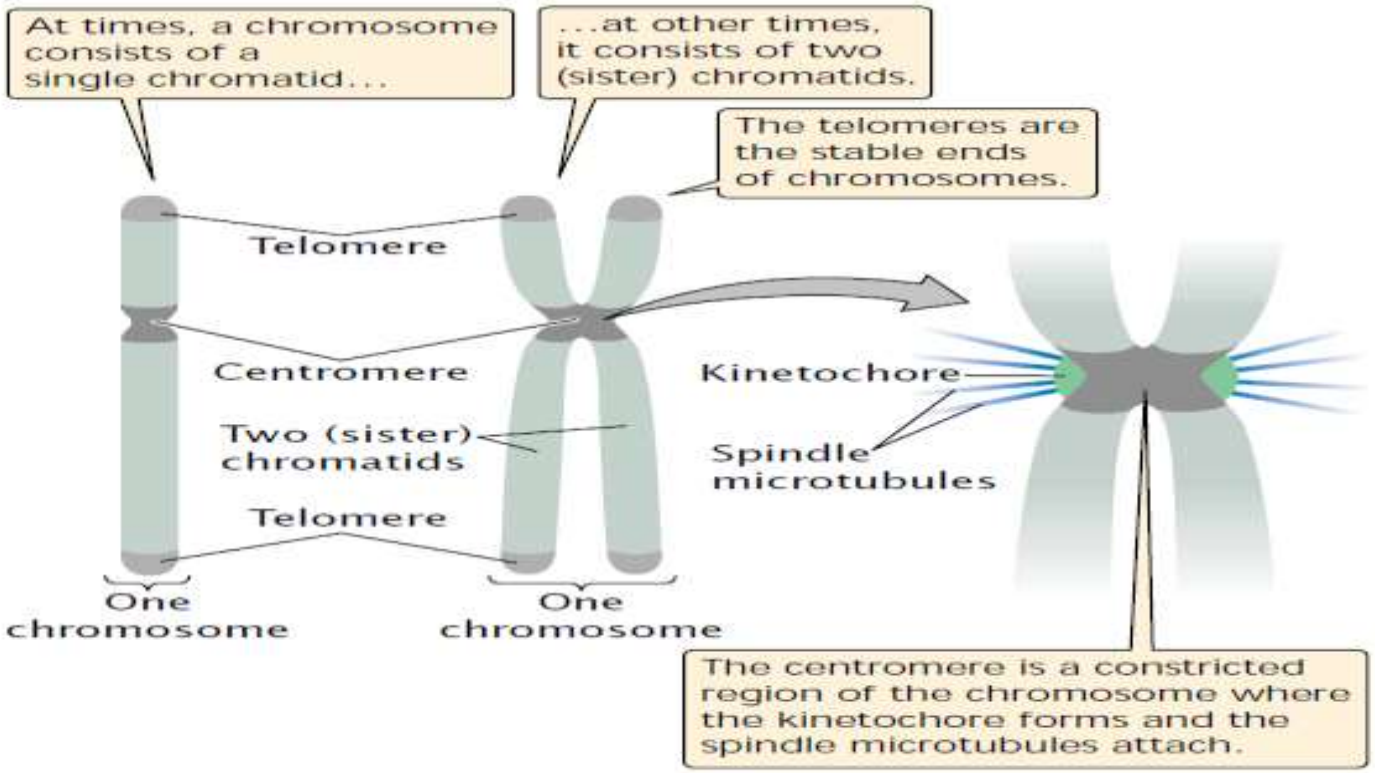
- Remain condensed throughout cell cycle and development
- Contains highly repetitive sequence
- Found at pericentromeric, Telomeric and loci along the chromosome
- Transcriptionally inactive
- Replicates late
- Reduced frequency of recombination

Facultative

- Regions of euchromatin converted to heterochromatin state
- Has potential for gene expression at some point in development
- Either can be condensed/ decondensed

Chromatin DNA molecule has 3 specific nucleotide sequence

- Centromere
- Telomere
- Origin of replication


At times, a chromosome consists of a single chromatid...
...at other times, it consists of two (sister) chromatids.
The telomeres are the stable ends of chromosomes.
Telomere
Centromere
Two (sister) chromatids
Telomere
One chromosome
One chromosome
Kinetochore
Spindle microtubules
The centromere is a constricted region of the chromosome where the kinetochore forms and the spindle microtubules attach.

2.7 **Structure of a eukaryotic chromosome.**

Source: Benjamin A Pierce, Genetics, A conceptual approach, Page no. 22,23

Centromere

- Constricted region of eukaryotic chromatin
- Kinetochore is assembled and sister chromatids are held together
- Not located exactly in a center of chromosome
- Regions on either side of Centromere- Chromosome's arms
- Kinetochore associated with Centromere- is a complex of proteins
- Where spindle fibers attach to chromosome during mitosis/ meiosis help in proper segregation of sister chromatids/ homologous chromosomes
- No defined DNA sequence
- Consists of large arrays of tandemly repeated DNA sequence
- Chromosomes classified based on location of Centromere:
  - Metacentric
  - Sub metacentric
  - Acrocentric
  - Telocentric
- In interphase stage-chromatin replicates formation of 2 copies of each chromatin
- Cell's M phase-Condensation of chromatin
- Formation of metaphase chromosome consisting of 2 identical sister chromatids
- Held together at Centromere(seen as constricted chromosomal region)- cohesin (protein links sister chromatids+ Centromere)

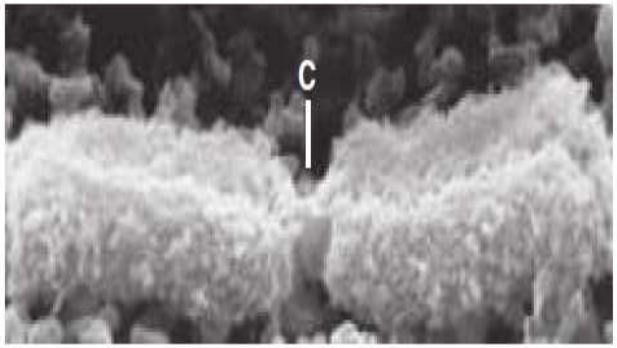
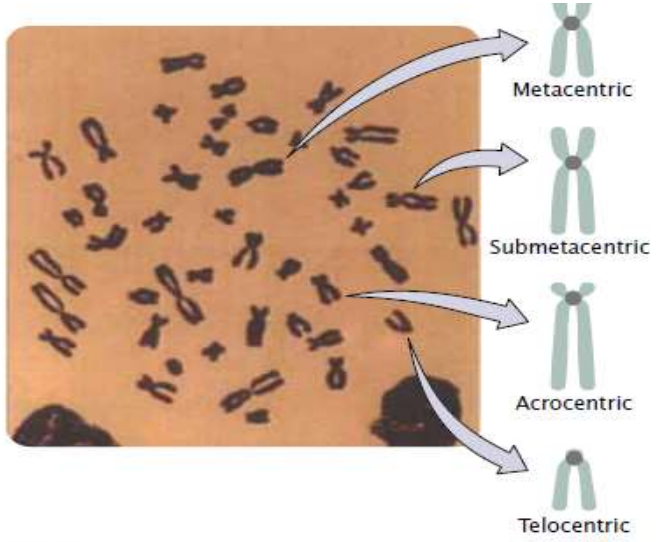


FIGURE 12.26 Each mitotic chromosome has a centromere whose site is marked by a distinct indentation. Scanning electron micrograph of a mitotic chromosome. The centromere (C) contains highly repetitive DNA sequences (satellite DNA) and a protein-containing structure called the kinetochore that serves as a site for the attachment of spindle microtubules during mitosis and meiosis (discussed in Chapter 14).

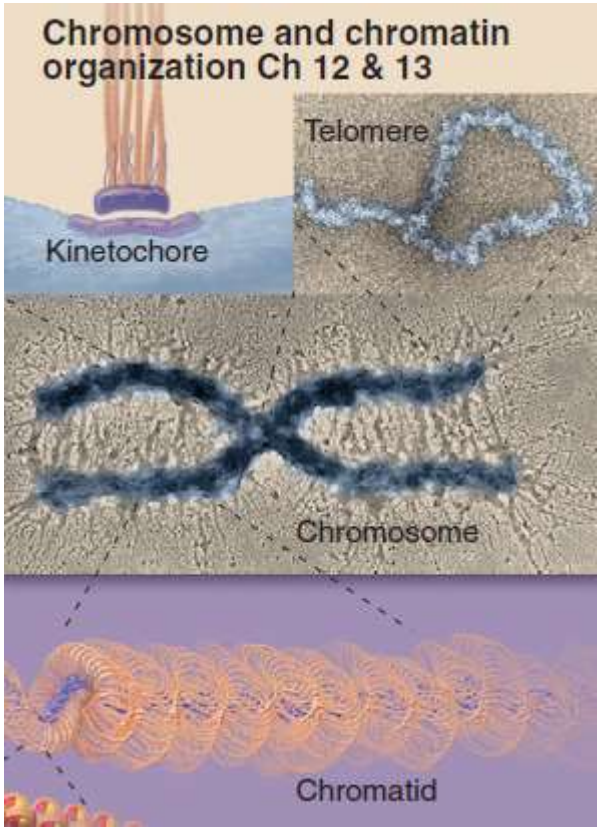
SOURCE: From Jerome B. Rattner, *Bioess.* 13:51, 1991, © 1991. This material is used by permission of John Wiley & Sons.



2.8 **Eukaryotic chromosomes exist in four major types.** (L. Lisco, D. W. Fawcett/Visuals Unlimited.)

Telomere

- Specialized structure
- Caps the end of eukaryotic chromosomes
- maintains structural integrity of a chromosomes
- Ensures complete replication of extreme ends of chromosome
- Consists of long array of short and tandemly repeated sequences
- 100-1000repeats
- Unusual property of Telomeric sequence is:
- Presence of G rich single strand 3’ overhang (50-300 nucleotides)
- G rich sequence is generated because there is limited degradation of C- rich complementary strand
- Sequence of telomeres has been highly conserved in evolution
- 2 sequence- specific DNA binding proteins- Telomeric repeat binding factor (TRF1) and Telomeric repeat binding factor 2 (TRF2) bind directly with Telomeric sequences, which in turn interact with a larger proteins
- Stability of chromosome ends- telomere
- Due to novel structure- “ t loop” found at ends of mammalian telomere
- Loops form at the ends of chromosomes under the direction of protein TRF2 which causes the 3’ end of chromosome to loop around and interdigitate into double helix, forming loop



Source: Thomas D pollard, William C Earnshaw, (2008), Cell biology, Saunders Elsevier, 2nd edition, Page no.212

Origin of replication

- Chromatins are multirepliconic
- Having more than one replicon (a replication unit)
- Individual replicons are relatively small, 40-100 kb
- Each replicon has its own origin of replication
- One chromatin contains multiple origin of replication
- In humans, chromatins have about 30000 origin of replication with each chromatin containing several hundreds

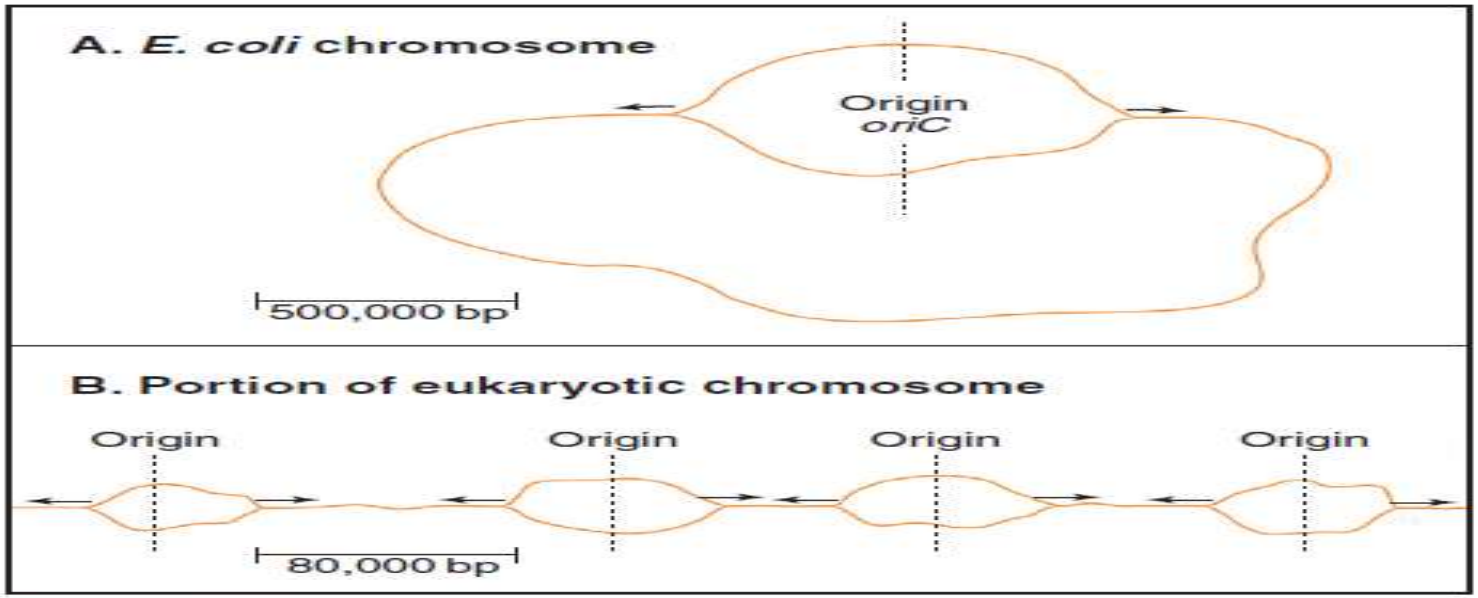


Figure 42-3 **A**, The *E. coli* chromosome is a simple replicon with a single origin of replication. In cells, this chromosome has a complex, highly supercoiled structure. **B**, Eukaryotic chromosomes have multiple origins of replication.

Source: Thomas D pollard, William C Earnshaw, (2008), Cell biology, Saunders Elsevier, 2nd edition, Page no.784

Rashmi M G

Chromosomal organization

Rashmi M G

Packaging of DNA into chromosomes

- In G˚ phase and Interphase- DNA exists as nucleoprotein complex- Chromatin
- In M phase- converted into more condensed form –chromosomes
- In Anaphase- single enormously long linear DNA molecule+ protein that fold and pack the fine DNA thread into more compact structure
- Total mass of histone in chromatin = DNA
- Average human chromosome- 5cm of DNA
- chromosome< DNA molecules (based on size)

Proteins that bind DNA to form eukaryotic chromosomes are divided into 2 classes

Histone chromosomal proteins
- Basic proteins
- Synthesized during S phase

Replication dependent histones
Ex. H1, H2A, H2B, H3, H4

Replication independent histones
Ex. H3.3

Non histone chromosomal proteins
- Acidic proteins
- Synthesized throughout the cell cycle
- Ex. Topoisomerase II

Histones

- Are responsible for first and most basic level of chromosome organization- The nucleosome (a basic structural unit of chromatin)
- Nucleosome contains, histone octamer(2 copies of H2A, H2B, H3 and H4)
- $(H3\text{-}H4)_2$ tetramer flanked by 2 H2A-H2B dimers – Core histones (rich in lysine and arginine; Positive charged amino acids which can effectively neutralize the negatively charged DNA backbone
- These 8 proteins form a barrel shaped core octamer, with the DNA wound around the outside (wrapped 1.7 turns in left handed coil)

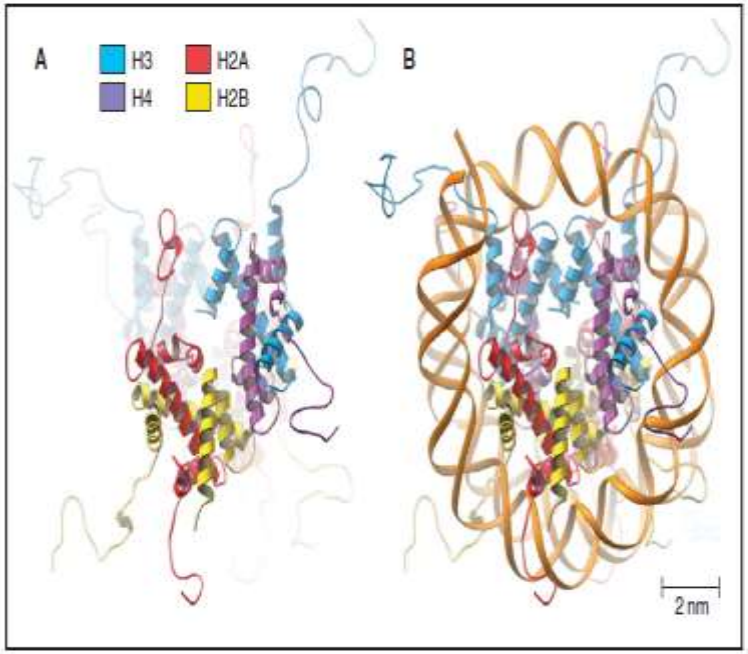
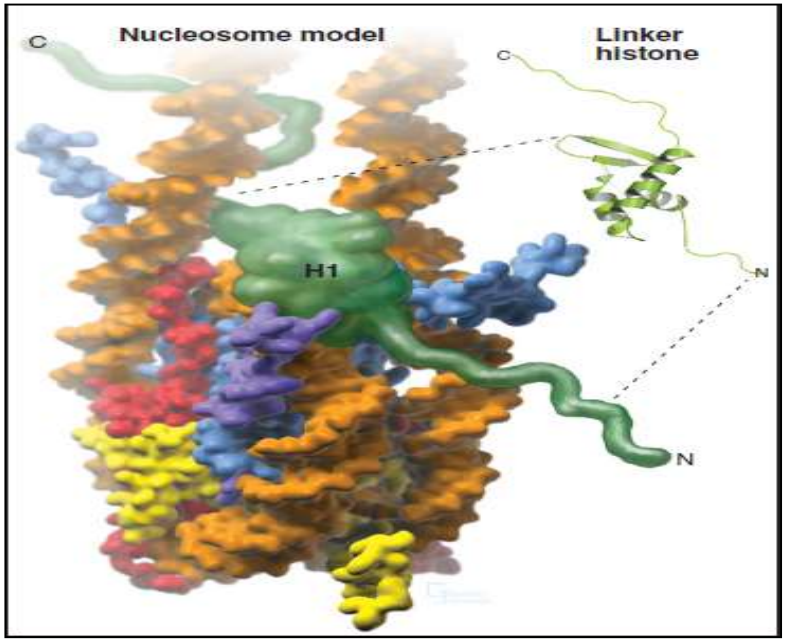


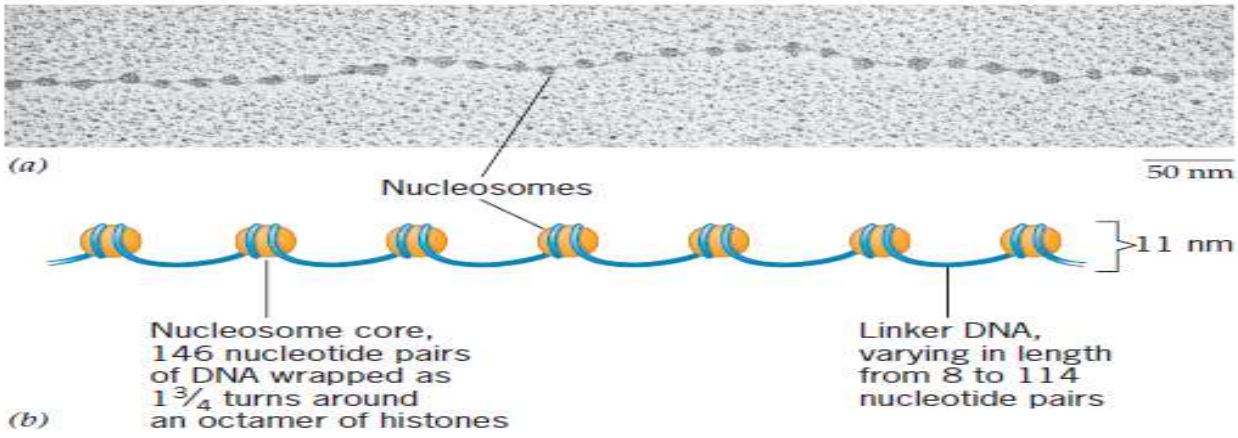
Figure 13-2 SECONDARY STRUCTURE OF THE HISTONES WITHIN THE CORE PARTICLE. **A,** A ribbon diagram shows that each histone protein in the octameric core of the nucleosome has a characteristic α-helical structure (the histone fold). Portions of the flexible N-terminal portions of the histones, which have a critical role in regulating chromatin structure, did not occupy a unique location in the crystal and do not appear in this structure. **B,** The histone octamer surrounded by one of the two turns of DNA. (PDB file: 1KX5. Modified from Luger K, Mäder AW, Richmond RK, et al: Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389:251–260, 1997.)



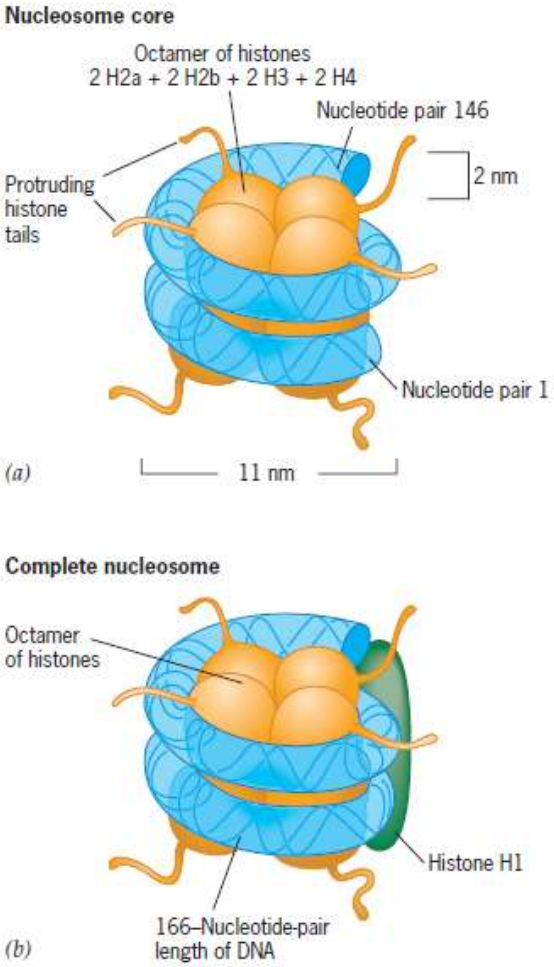
Source: Thomas D pollard, William C Earnshaw, (2008), Cell biology, Saunders Elsevier, 2nd edition, Page no.232,235

Nucleosome

- Provide 1st level of DNA packing
- Gives basic bead- on- string structure
- On treatment with micrococcal; nuclease initially cleaves the DNA between nucleosomes and produce mono nucleosomes
- Each nucleosome is separated by 50-70 bp of linker DNA giving repeat length of 190-220bp
- H1 termed as linker histone, is associated with linker DNA and may lie at the point where DNA enters and leaves the nucleosome
- Single linker histone is attached to each nucleosome- Chromatosome


(a)
50 nm
Nucleosomes
11 nm
Nucleosome core, 146 nucleotide pairs of DNA wrapped as 1¾ turns around an octamer of histones
Linker DNA, varying in length from 8 to 114 nucleotide pairs
(b)

■ **FIGURE 9.17** Electron micrograph (*a*) and low-resolution diagram (*b*) of the beads-on-a-string nucleosome substructure of chromatin isolated from interphase nuclei. *In vivo*, the DNA linkers are probably wound between the nucleosomes forming a condensed 11-nm fiber.

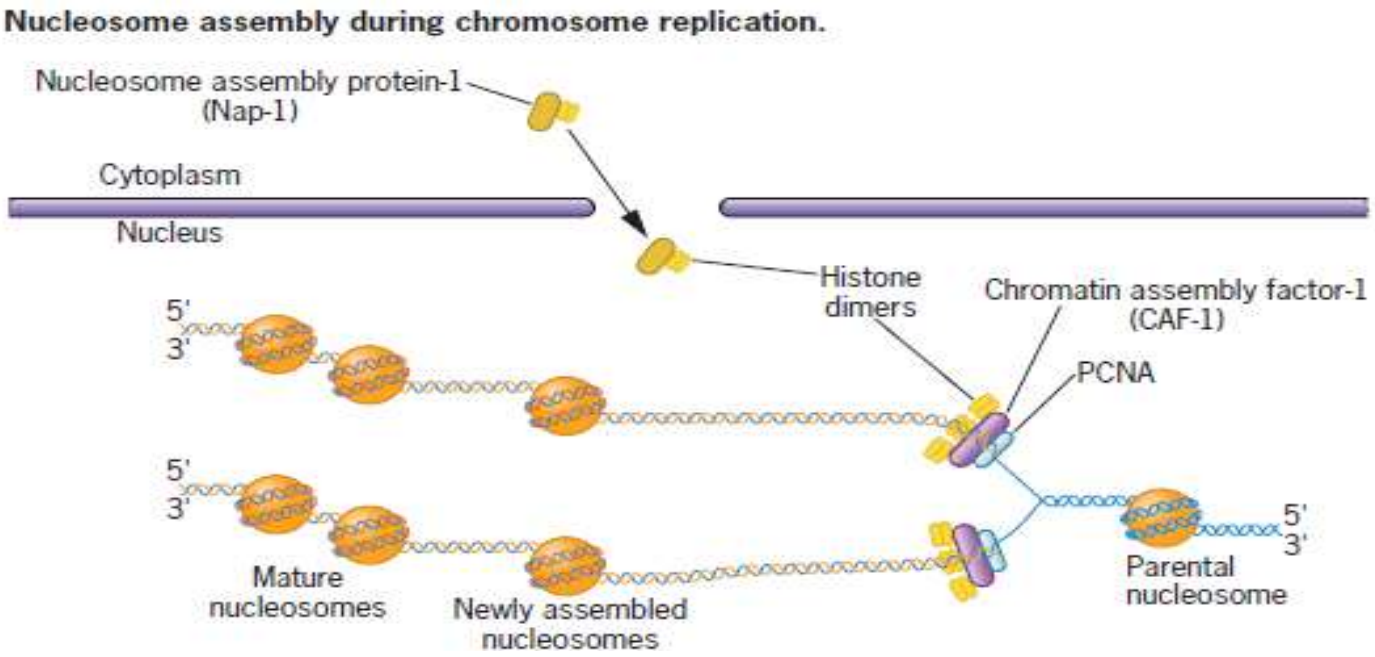


■ **FIGURE 9.18** Diagrams of the gross structure of (*a*) the nucleosome core and (*b*) the complete nucleosome. The nucleosome core contains 146 nucleotide pairs wound as 1.65 turns of negatively supercoiled DNA around an octamer of histones—two molecules each of histones H2a, H2b, H3, and H4. The complete nucleosome contains 166 nucleotide pairs that form almost two superhelical turns of DNA around the histone octamer. One molecule of histone H1 is thought to stabilize the complete nucleosome.

Source: Snustad, Simmons (2012), Principles of genetics, John Wiley& Sons, Inc, Sixth edition, Page no. 208,209

Nucleosome assembly

- In nucleosome assembly histone binding to DNA is thought to occur in 2 step process:
- A tetramer / 2 dimer of histones H3 and H4 are first deposited onto DNA
- 2 dimers of histones H2A and H2B are deposited to yield nucleosome core
- Nucleosome assembly proteins termed histone chaperones are crucial for incorporation of histones into nucleosomes
- Histones are closely escorted by histone chaperones from their point of synthesis upto their delivery site
- Many histone chaperones are identifies: nucleoplasmin, ASF1(anti-silencing function 1), HIR protein (Histone regulator), CAF1 (chromatin assembly factor 1) and NAP1 (Nucleosome assembly protein 1)



Source: Snustad, Simmons (2012), Principles of genetics, John Wiley& Sons, Inc, Sixth edition, Page no. 247

30 nm fiber

- 2nd level of organization
- Arrangement of nucleosome- in a fiber of diameter 30nm
- 2 popular models(based on in vitro data):

Solenoid model

- One start helix model
- Nucleosome follow each other along the same helical path
- Interaction between the nucleosomes occurs sequentially
- Characterized by interaction between consecutive nucleosome
- (n, n+1, n+2 so on)
- Interaction between the histone core(1,2,3 and so on)

Zig zag model

- Two strand helix model
- Two strands containing stacked nucleosome are wound into a left handed two strand helix
- In each strand nucleosomes are aligned with each other like a stacked coin
- Interaction between alternate nucleosomes
- (n , n+2, and so on)
- Interaction between histone core (1 and 3, 2 and 4 and so on)

Rashmi M G

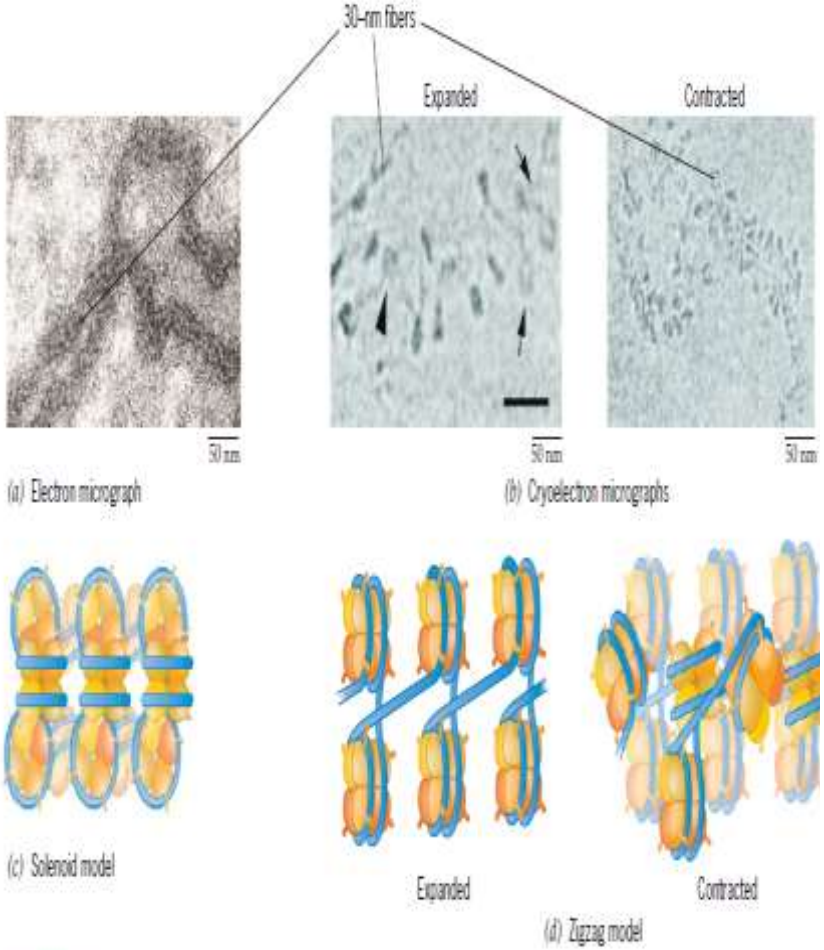


FIGURE 9.21 Electron micrograph (a) and cryoelectron micrographs (b) of the 30-nm chromatin fibers in eukaryotic chromosomes. The structure of 30-nm chromatin fibers seems to vary based on the procedures used to isolate and photograph them. (c) According to one popular model, the 30-nm fiber is produced by coiling the 11-nm nucleosome fiber into a solenoid structure with six nucleosomes per turn. (d) However, when chromatin is visualized after cryopreservation (quick freezing) without fixation, it exhibits a zigzag structure whose density—expanded versus contracted—varies with ionic strength and with chemical modifications of the histone molecules.

Source: Snustad, Simmons (2012), Principles of genetics, John Wiley& Sons, Inc, Sixth edition, Page no. 211

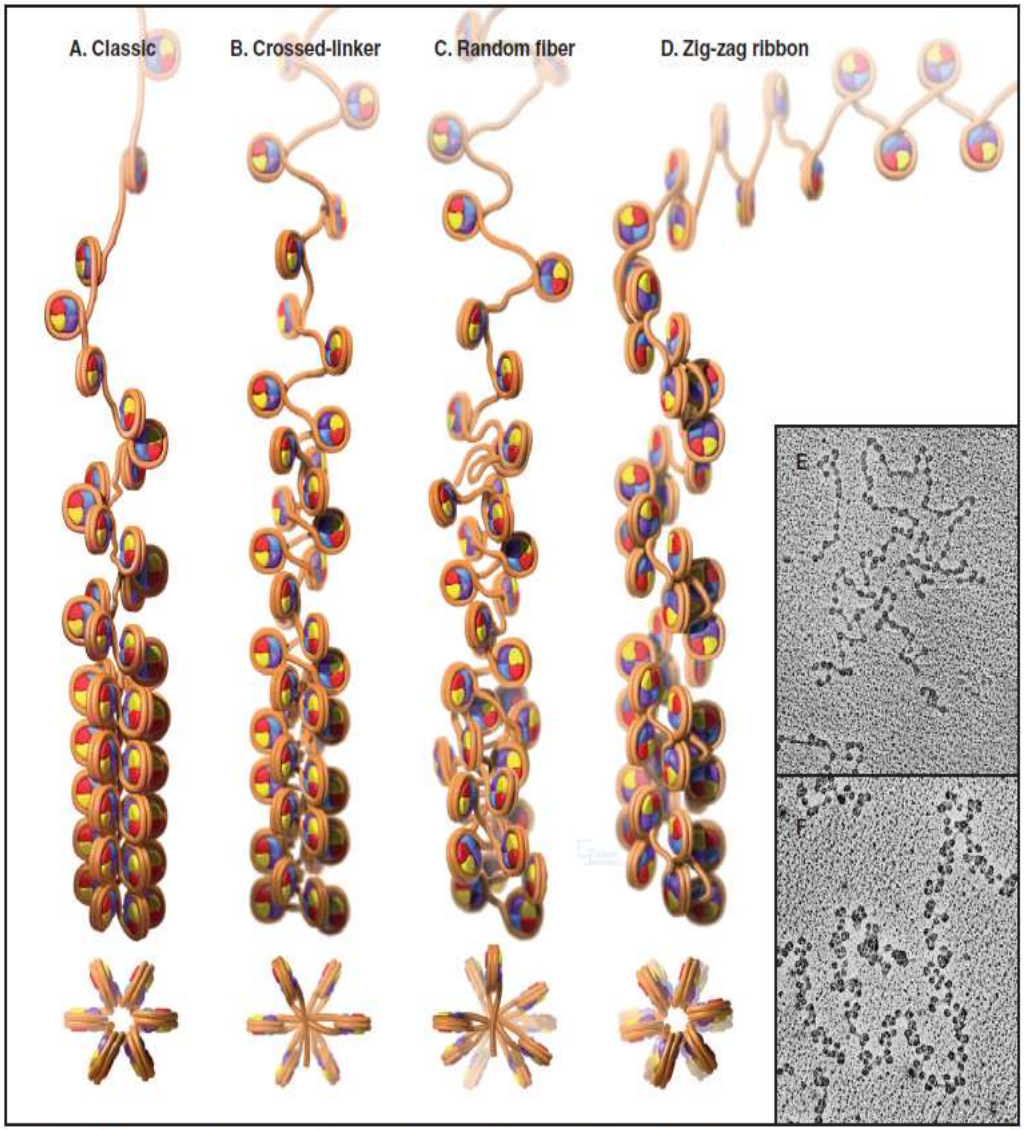


Figure 13-6 ALTHOUGH IT HAS BEEN STUDIED INTENSIVELY FOR 30 YEARS, THE STRUCTURE OF THE 30-NM CHROMATIN FIBER REMAINS CONTROVERSIAL. **A–D,** Various models of the 30-nm fiber: classic solenoid (a type of helix), crossed-linker solenoid, random fiber, and zig-zag ribbon. The classic solenoid model was originally favored, but it now appears that the crossed-linker and zig-zag ribbon are more likely. **E–F,** Histone H1 causes a compaction of the chromatin filament. **E,** Chromatin spread for electron microscopy following removal of histone H1: the 10-nm fiber or so-called beads on a string. **F,** A similar preparation with H1 present: the 30-nm fiber. (E–F, From Thoma F, Koller T: Influence of histone H1 on chromatin structure. Cell 12:101–107, 1977.)

Source: Thomas D pollard, William C Earnshaw, (2008), Cell biology, Saunders Elsevier, 2nd edition, Page no. 236

Rashmi M G

Metaphase chromosome

- In interphase- 30nm fiber- major type of chromatin in nucleus
- When nucleus divides- DNA adopts more compact form of packaging
- Highly condensed- metaphase chromosome
- DNA compact
- 1/10000 of its stretch out length
- Loops of 30 nm chromatin fiber containing 20-100kb of DNA/loop are attached to a central **scaffold**
- Scaffold consists of several proteins:
  - (non histone acidic proteins) Topoisomerase II,
  - chromatin proteins
  - condensin proteins
- **Topoisomerase** II- Enzyme having ability to pass one DNA double helix through another by cutting and sealing both strands of DNA
- **Chromatin proteins-** binds to AT rich sequence chromatin loops attached by stretched of several hundred base pairs of highly AT rich (>70%) DNA- Scaffold Attachment Region (SAR)
- **Condensin proteins-** Maintains condensed chromosome structure, made up of SMS protein (structural maintenance of chromosome protein)
- SMC proteins are core components of Cohesins and Condensins
- **Cohesins-** links together the sister chromatids immediately after replication and keeps them together
- **Condensins-** make chromosome highly condense

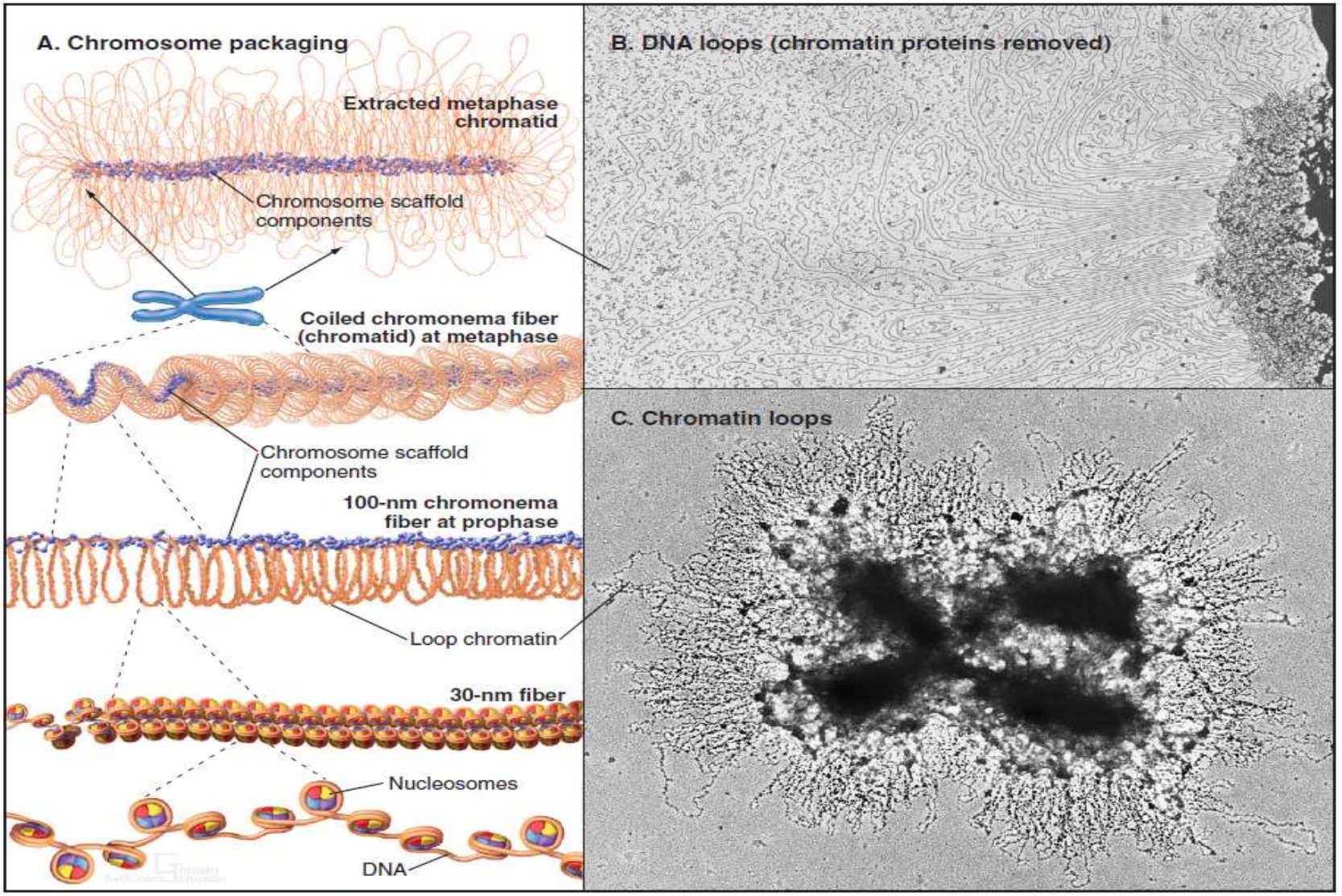
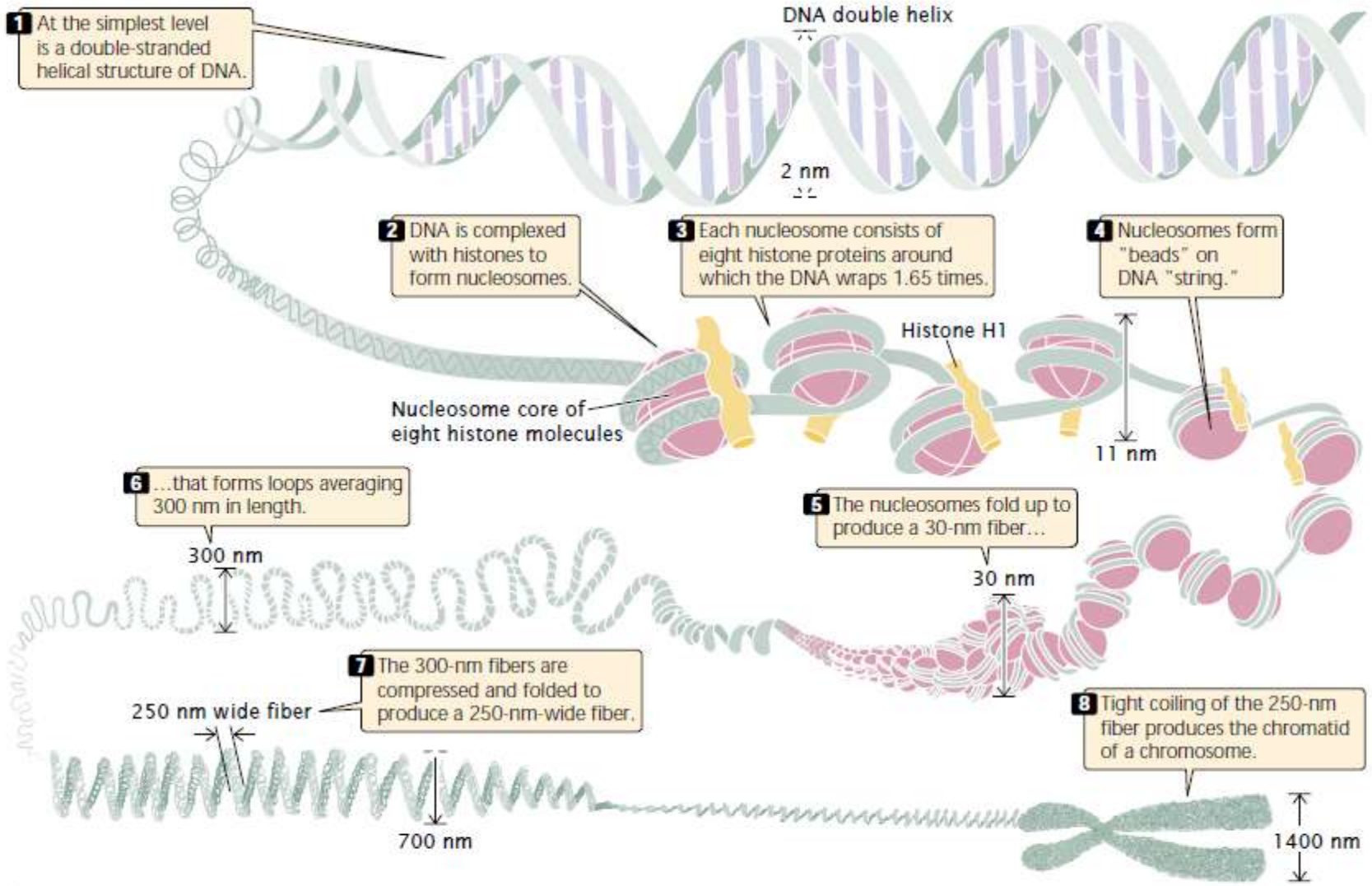


Figure 13-16 CURRENT MODELS OF MITOTIC CHROMOSOME STRUCTURE. **A,** Filament of nucleosomes, 30-nm fiber, chromonema fiber, coiled chromonema fiber. Nonhistone proteins complexes *(red dots)* bind and end up concentrated along the central axis of the chromatid arm. Cross-links between these complexes create the chromosome scaffold. When chromosomes are swollen or extracted, the scaffold remains compact, and loops of chromatin or DNA radiate out from it. **B,** DNA loops seen in a human mitotic chromosome from which the histones had been removed. **C,** Human chromosome showing loop domains. (B, From Paulson JR, Laemmli UK: The structure of histone-depleted chromosomes. Cell 12:817–828, 1977. C, Courtesy of William C. Earnshaw.)

Source: Thomas D pollard, William C Earnshaw, (2008), Cell biology, Saunders Elsevier, 2nd edition, Page no.245


1 At the simplest level is a double-stranded helical structure of DNA.
DNA double helix
2 nm
2 DNA is complexed with histones to form nucleosomes.
3 Each nucleosome consists of eight histone proteins around which the DNA wraps 1.65 times.
4 Nucleosomes form "beads" on DNA "string."
Histone H1
Nucleosome core of eight histone molecules
11 nm
6 ...that forms loops averaging 300 nm in length.
300 nm
5 The nucleosomes fold up to produce a 30-nm fiber...
30 nm
7 The 300-nm fibers are compressed and folded to produce a 250-nm-wide fiber.
250 nm wide fiber
8 Tight coiling of the 250-nm fiber produces the chromatid of a chromosome.
700 nm
1400 nm

11.5 Chromatin has a highly complex structure with several levels of organization.

Source: Benjamin A Pierce, Genetics, A conceptual approach, Page no. 295

Rashmi M G

Histone modification

Rashmi M G

Histone modification

All of histone protein are chemically modified

Leads to structural changes occurs at the time of replication and transcription

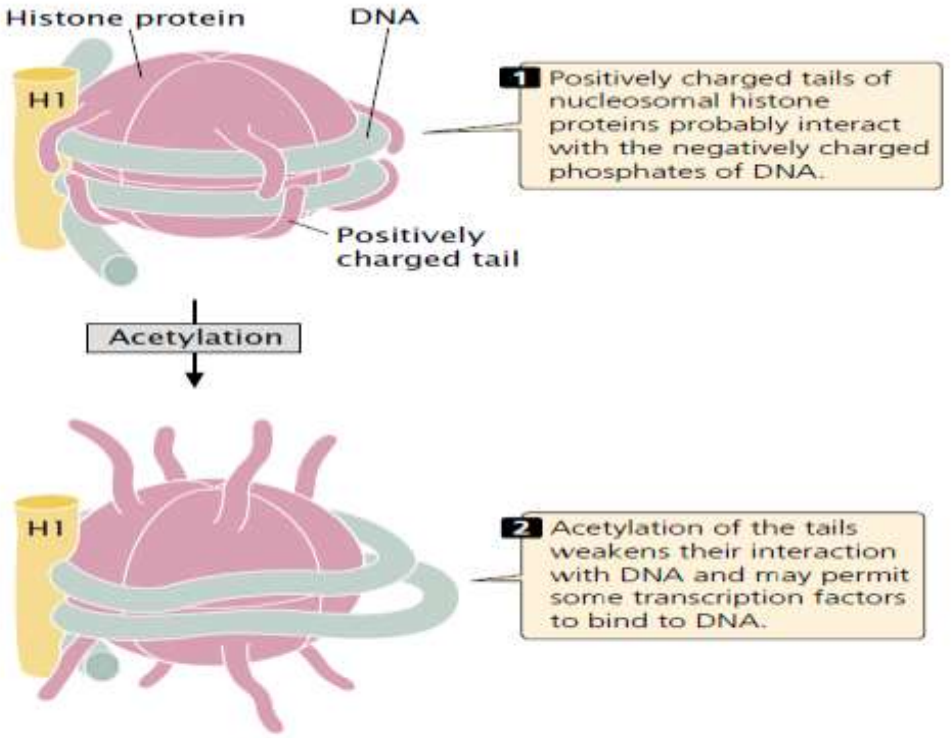
8 types of modifications found in histones:

1. Acetylation
2. Methylation
3. Phosphorylation
4. Sumoylation
5. Ubiquitylation
6. ADP Ribosylation

Type of modification	Residues modified	Function regulated
Acetylation	Lysine	Transcription, repair, condensation
Methylation	Lysine, arginine	Transcription, repair
Phosphorylation	Serine, threonine	Transcription, repair, condensation
Ubiquitylation	Lysine	Transcription, repair
Sumoylation	Lysine	Transcription
ADP Ribosylation	Glutamate	Transcription

Acetylation:

- Histone protein making up nucleosome core contain flexible amino terminus of 11-37 residues
- extending from fixed structure
- These N termini are called Histone tails
- Acetyl group added to lysine amino acids in histone tail of each of core molecules
- Key enzyme- Histone Acetylase Transferase HAT(commonly called as histone Acetylase)


Histone protein
DNA
H1
1 Positively charged tails of nucleosomal histone proteins probably interact with the negatively charged phosphates of DNA.
Positively charged tail
Acetylation
H1
2 Acetylation of the tails weakens their interaction with DNA and may permit some transcription factors to bind to DNA.

16.20 **The acetylation of histone proteins alters chromatin structure and permits some transcription factors to bind to DNA.**

Source: Benjamin A Pierce, Genetics, A conceptual approach, Page no. 458

Effects:

- Reduces affinity of histone with DNA (removes + charge from lysine)
- Reduce interaction between individual nucleosome
- Leads to formation of 30 nm chromatin fiber
- Reversible
- Enzyme Histone Deacetylases (HDACS) CATALYZE THE REMOVAL of acetyl group (deacetylation)
- Histones:
- Heterochromatin- deacetylated
- Euchromatin- Acetylated- creates docking site for binding of other proteins

Methylation

- Lysine and arginine amino acids at N terminus of histone undergo methylation
- Key enzyme- Methyl transferase
- Methylation:
  - At lysine,
    - At 9th position (H3K9)- gene silencing chromatin condensation
    - At 4th position (H3K4)- transcription activation

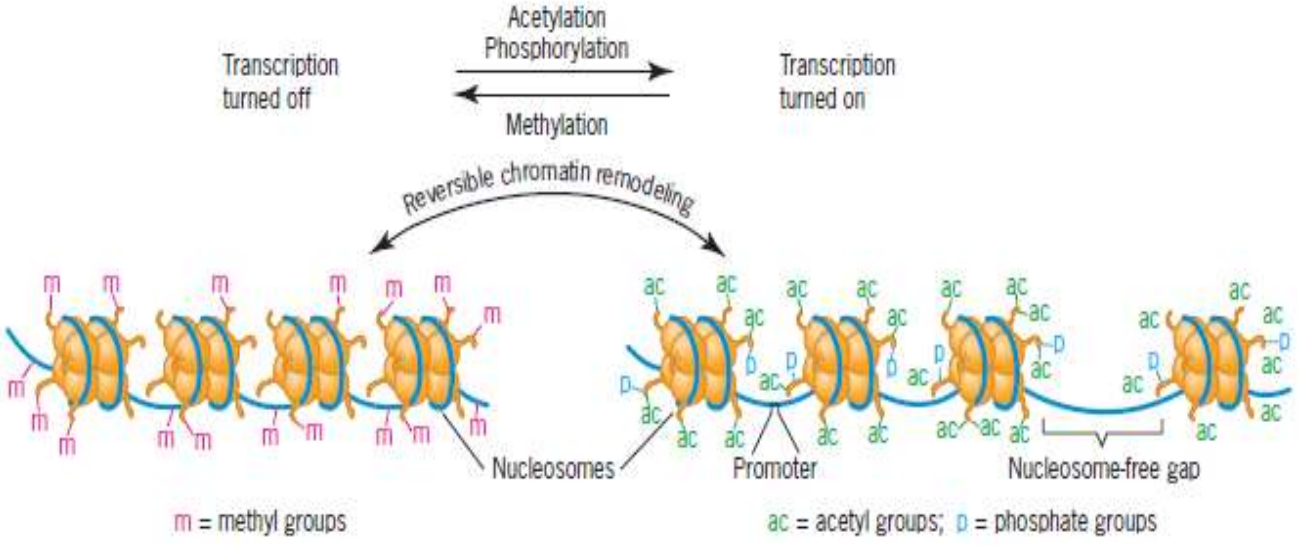


FIGURE 2 A schematic overview of the effects of (1) methylation of DNA and histones, (2) acetylation of histones, and (3) phosphorylation of histones on chromatin remodeling and transcription.

Effects:

- Role in heterochromatin formation
- Chromosome X inactivation
- Genome imprinting
- Transcriptional regulation

Phosphorylation

- Influence chromatin structure
- Impact on cellular activity
- Phosphorylation of H3 and H1 –formation of metaphase chromosome

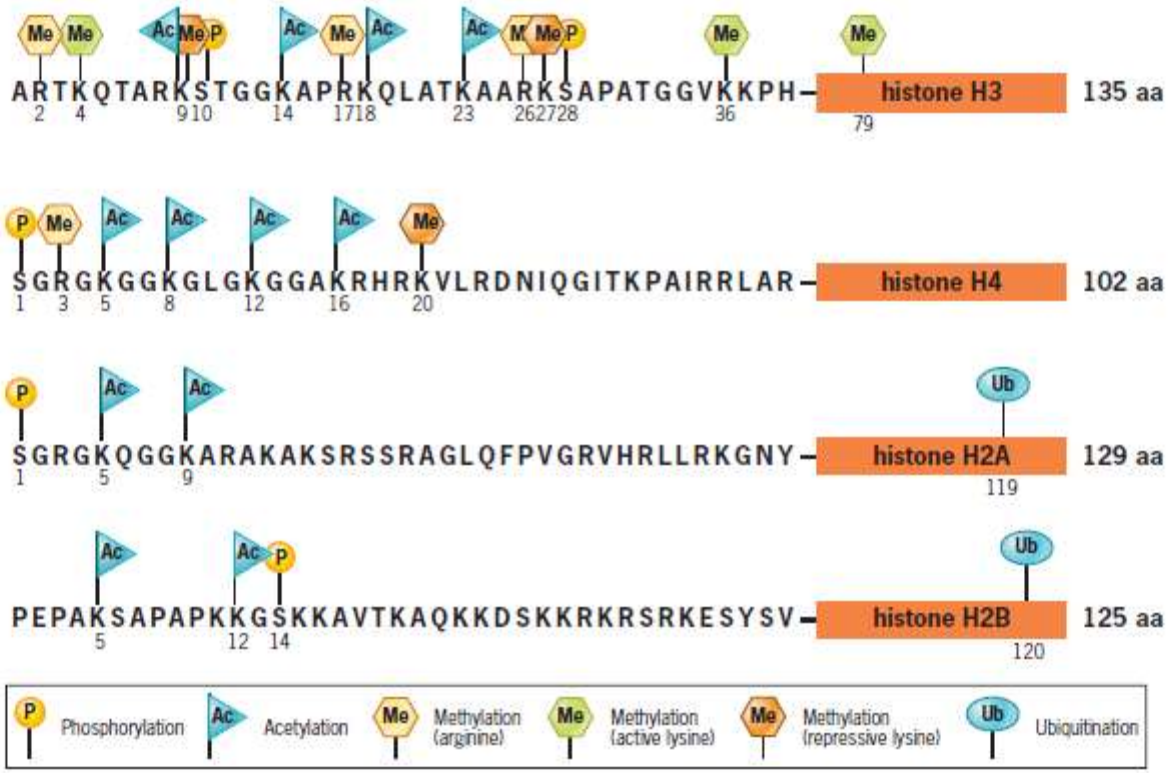


FIGURE 12.18 Histone modifications and the histone code. The amino terminal tails from histone proteins that extend out past the DNA in nucleosomes can be enzymatically modified by the covalent addition of methyl, acetyl, and phosphate groups. Methyl groups are added to either lysine (K) or arginine (R) residues, acetyl groups to lysine residues, and phosphate groups to serine (S) residues. The small protein ubiquitin can be added to one of the lysine residues in the core (rather than the tail) of H2A and H2B. Each lysine residue can have one, two, or three added methyl groups, and each arginine residue can have one or two added methyl groups. These modifications affect the affinity of the histone for interacting proteins that control the transcriptional activity of chromatin, which has led to the concept of a histone code. It is widely found that acetylation of lysines leads to transcriptional activation. The effects of methylation depend strongly on which of these residues is modified. For example, methylation of lysine 9 of histone H3 (i.e., H3K9) is typically present in heterochromatin and associated with transcriptional repression, as discussed in the text. Methylation of H3K27 and H4K20 is also strongly associated with transcriptional repression, whereas methylation of H3K4 and H3K36 is associated with transcriptional activation. Just as there are enzymes that catalyze each of these modifications, there are also enzymes (deacetylases, demethylases, phosphatases, and deubiquitinases) that specifically remove them.

SOURCE: From C. David Allis et al., *Epigenetics*, Figure 3.6, page 31, Copyright 2007. Reprinted with permission from Cold Spring Harbor Press.

Source: Janet Iwasa, Wallace Marshall (2016), Karp's cell and molecular biology, 8th edition, Page no.501